SEC

AD-A174 571 ORT DOCUMENTATION PAGE

(2)

1a. U	1b. RESTRICTIVE MARKINGS NA
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Cornell University #E60-8367	5. MONITORING ORGANIZATION NA
6a. NAME OF PERFORMING ORGANIZATION Cornell University; 6b. OFFICE SYMBOL (If applicable) NA	7a. NAME OF MONITORING ORG Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Ithaca, New York 14853-5201 Olin Hall of Chemical Engineering	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research; 8b. OFFICE SYMBOL (If applicable) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-85-K-0580
8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61153N; PROJECT NO. RR04106; TASK NO.; WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
(U) Development of Cell Models as a Basis for Bioreactor Design for Genetically Modified Bacteria

12 PERSONAL AUTHOR(S)
Michael L. Shuler

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 7/85 TO 7/86	October 30, 1986	

16. SUPPLEMENTARY NOTATION
Project Sponsored Jointly by Materials Division and Molecular Biology

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Mathematical Models; Genetic Instability; *Escherichia coli*; Plasmids; Bioreactors

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The purpose of the project is the development of a mathematical model of the bacterium, *Escherichia coli*, modified by the insertion of plasmids containing foreign DNA. Such a model provides a basis to simulate on the computer the response of a culture to a variety of environmental stimuli. The results of such simulations can be verified experimentally. This model is formulated to allow the investigator to probe hypotheses concerning cellular regulation and biochemistry. This ability is particularly important with respect to genetically modified cells where the insertion of foreign DNA and expression of foreign genes interacts strongly with host cell physiology in complex and often unpredictable ways. The model should provide, by the completion of the project, useful guidance on how to "best" construct a plasmid for insertion into *E. coli*.

Not only should the model suggest optimal strategies for construction of strains but should provide the engineer with a tool to better design large-scale systems utilizing cells with recombinant DNA. Major problems in the large scale utilization of plasmid

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	U
22a NAME OF RESPONSIBLE INDIVIDUAL Michael L. Shuler	22b. TELEPHONE (Include Area Code) 202/696-4760; 22c. OFFICE SYMBOL ONR

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

containing cells are genetic instability and maintenance of high levels of gene expression.

The starting point for this project has been a class of models developed previously by our group for *Escherichia coli*. The base model is that for an individual cell. All cellular components are distributed into 20 model components (e.g. chromosonal DNA, stable RNA, m-RNA, etc.). Population models are constructed from an ensemble of single-cell models. Population models account for distribution of cellular capabilities among a population. The base single-cell model responds explicitly to changes in glucose or ammonium ion concentrations in the medium and mimics growth under either fully aerobic or anaerobic conditions.

During the first year of the project, we have successfully introduced a model for the control of replication of plasmids with the ColE1 origin of replication. Model predictions of copy number compare well to reported experimental results. If we force the model to make a given amount of plasmid encoded protein, we can then predict genetic instability for a culture producing the same level of plasmid-encoded protein. The model prediction and experimental results show that stability (in terms of number of generations before a revertant reaches 90% of the total population) is greater in faster growing than slow growing cultures. Further we have extended the model to the low growth rate regime ($< 0.15\ hr^{-1}$). This extension has allowed us to make predictions of the effect of growth rate on plasmid copy number. The slow growth region is of interest since immobilized cell reactors may be useful commercially and some other applications, such as *in situ* treatment of wastes may place cells under conditions of restricted growth. Extension of the model to include amino acids as substrates and an explicit model of the lac promoter is now partially complete.

Discussion of Results for Year 1

The purpose of the project is to develop a mathematical model for *Escherichia coli* carrying plasmids in which a high level of plasmid encoded protein synthesis can be induced. The model can then be used to guide strain construction and the development of bioreactors and appropriate operating strategies.

Our proposal to ONR was based on previous work we had done developing models for *Escherichia coli*. Early attempts to write models of populations with both chemical structure and segregation (i.e. recognition that a population contains a destruction of properties) failed. These attempts failed because the models were written from a population perspective and the resulting equations were intractable. We circumvented these problems by writing a detailed model of an individual cell which contained significant chemical structure and then constructed a population model using a finite representation technique. This approach results in a mathematically tractable model that predicts the dynamic (i.e. transient) response of a population to perturbations in environmental variables such as substrate concentration (glucose or ammonium ion) or flow rate in a chemostat. Review articles describing this modeling approach have appeared during this last year (Shuler, 1985. Chem. Eng. Commun. *36*:161; and Shuler, 1985. In *Comprehensive Biotechnology Vol. 1*, p. 119) and reprints are included in this report. Although these articles appeared during the period of ONR support, they were written prior to the initiation of the ONR grant.

We have successfully introduced a model for the replication of ColE1 type plasmids into the base single-cell model of *E. coli*. The results are described in the manuscript entitled "Mathematical Model for the Control of ColE1 Type Plasmid Replication" which is currently in press for publication in *Plasmid* (copy attached).

RNAII is a transcript that combines with the plasmid origin of replication to form a preprimer for replication. If RNAI binds with RNAII before it reaches the origin, no replication can occur. ROM is a plasmid-encoded protein that enhances RNAI binding. Using binding-constants for RNAI-RNAII measured by Tomizawa and colleagues for various mutations in the absence and presence of the ROM protein, we are able to predict copy numbers similar to that measured by Tomizawa and colleagues. The model even predicts correctly those cases where copy number exceeds 100 which former models have been unable to do. Using the model we were able to test three hypotheses about the mechanism of interaction of RNAI and the Rom protein. The model favorably supports the mechanism proposed by Tomizawa and colleagues concerning the nature of RNAI-RNAII interactions and that the ROM protein increases the binding rate between the two RNA species. The hypothesis that the interactions of RNAI-RNAII increases the susceptibility of RNAII to the action of endonucleases is not a plausible mechanism.

The production of high levels of plasmid-encoded proteins leads to a "metabolic burden" on host cells. Revertants that lose the plasmid through missegregation of the plasmid have a distinct growth advantage over plasmid-containing cells. Consequently nonproductive revertant cells can

displace plasmid-containing cells or what is often called "genetic instability". By using our model with a host cell containing plasmid we can make reasonable predictions of genetic instability. These results are discussed in the enclosed manuscript, "A Mathematical Model for Predicting Copy Number and Genetic Stability in *Escherichia coli*" which has been accepted for publication in *Biotechnology and Bioengineering*. The model over-predicts the degree of stability of the culture slightly. The model only recognizes segregational plasmid instability. Since structural instability or the formation of multimers is possible in the culture used for experimental verification, a slight overprediction of stability is understandable. Both model and experiment show that the system is more stable at $\mu = 0.7\ hr^{-1}$ than at $\mu = 0.3\ hr^{-1}$.

For many practical systems it may be desirable to operate for a period of time under fairly severe nutrient limitations. Cyclic reactors, immobilized cell systems, or *in situ* treatment of hazardous wastes are examples. It may also be of importance to understand what steps become growth rate controlling under a variety of environmental circumstances. In a manuscript entitled, "Growth Behavior and Prediction of Copy Number and Retention of ColE1 Type Plasmids in *E. coli* Under Slow Growth Conditions" submitted for publication in a volume of *Annals NY Acad. Sci.* we report on some of our results. Under severe glucose limitations the cell is starved for energy - primarily needed to maintain membrane energization. At less severe conditions glucose is limiting primarily as a supplier of carbon for cellular material. If the transcription rates for RNAI and RNAII decrease at the same rate as the decrease in overall transcription for glucose limitation, then copy number is relatively constant for dilution rates greater than $0.3\ hr^{-1}$ but increase dramatically at very low growth rates.

We are currently working on extending the model to respond explicitly to the addition of amino acids since such additions are commonly used commercially. We have nearly completed work on a model for glutamine incorporation. We have also begun work (although we are in the formative stages) on inserting a mathematical model of the lac promoter into the plasmid-cell model. The model will allow us to predict the dynamic response of a culture to induction with IPTG and also to predict the level of overproduction of plasmid-encoded protein.

Chem. Eng. Commun. Vol. 36 pp. 161–189
0098-6445/85/3601–0161$25.00/0

INVITED REVIEW ON THE USE OF CHEMICALLY STRUCTURED MODELS FOR BIOREACTORS

M.L. SHULER

School of Chemical Engineering Cornell University Ithaca, New York 14853

(Received August 6, 1984; in final form December 11, 1984)

An individual cell is an immensely complicated self-regulated chemical reactor that can alter its biosynthetic machinery to meet the demands of a changing environment. The biochemical engineer must build a large macroscopic reactor to harness the cells for desirable chemical conversions. The design and control of such bioreactors would be facilitated with effective mathematical models of the response of the culture to changes in nutrients or other environmental variables. Because of the inherent internal plasticity of the cell, models must reflect the changing structure of the biomass. This paper reviews some examples of models which contain components representing various chemical fractions within the cell. The advantage of these models is their potential ability to predict the dynamic behavior of a cellular population. In addition such models are potential tools for testing hypotheses concerning cellular control mechanisms and consequently the development of more effective cell strains. Models of populations based on a finite-representation technique using an ensemble of chemically structured single-cell models are emphasized. These latter models are capable of accurate *a priori* prediction of bioreactors to perturbations in flow rates or feed concentrations. Models which aspire to the *a priori* quantitative prediction of cell population behavior must be sufficiently complex that shifts in growth-rate limiting processes can be taken into account; consequently a high-level of chemical structure will characterize the best models.
KEYWORDS Bioreactors, *Escherichia coli*, Structured mathematical models

INTRODUCTION

Biochemical reaction engineering differs significantly from traditional chemical reaction engineering due to the nature of the "catalytic reagent". A catalyst such as platinum on a support is relatively fixed in its catalytic properties and responds very slowly and passively to changes in its environment. A living cell is an independent chemical reactor; a very complex reactor with more than a thousand individual reactions operating under a highly sophisticated control system. Because of this control system the biosynthetic capabilities of a cell actively change in response to its environment. Thus the bioreactor engineer is faced with the difficult task of designing a macroscopic reactor to provide an optimal environment for a multitude of individual cellular reactors nested inside the macroscopic reactor. Optimal in this case implies maximum expression of a population's genetic potential for a given set of chemical transformations. This inherent plasticity of the biological catalyst differentiates biochemical reaction engineering from its traditional counterpart.

In traditional chemical reaction engineering mathematical models are important

to reactor design and control and also as tools to discern possible reaction mechanisms. Perhaps in bioreactors such models are of even greater importance because of the inherent complexity of biological systems and the difficulty in directly measuring important growth parameters. For example, in a traditional reactor the measurement of temperature and pressure often allow the engineer to determine reactor state. The state of a bioreactor cannot be measured so easily (e.g. temperature and pressure do not fix system properties).

However, we cannot usefully describe bioreactors with mathematical models unless the underlying physical reality is understood. The next section gives a very brief description of cellular systems.

HOW A CELL WORKS

From the engineer's point of view the key elements in a cell are enzymes. Enzymes are catalysts which generally have a high level of specificity and fairly rapid reaction rates at near ambient conditions (e.g. typical turnover numbers are in the range of 10^2 to 10^6 molecules of product/min/active site (Mahler and Cordes, 1966)). A simple bacterium such as the common intestinal organism *Escherichia coli* can produce over 1000 different enzymes (Watson, 1976). Each enzyme is a protein; proteins are polymers of amino acids with molecular weights in the range of about 10^4 to 10^6 daltons. Only the L-isomer of each amino acid can be incorporated into proteins. Twenty-one different amino acids are typically used. The sequence of amino acids gives the protein its primary structure. However, an enzyme can only function if it assumes the appropriate three-dimensional structure (e.g. secondary and tertiary structure) which depends on hydrogen bonds, disulfide bonds, and salt bonds as well as hydrophobic and hydrophilic interactions. Since these bonds are relatively weak, most enzymes retain their three-dimensional shape and activity over a relatively small range of temperature and pH.

Proteins (which can have a structural roles as well as being enzymes) constitute about 50 to 70% of the cell's dry weight. The "blueprint" for each protein is encoded on the cell's DNA or chromosome. The code for each protein can be transcribed to a messenger RNA molecule (m-RNA). The message is then translated into an actual protein molecule using ribosomes (which contain both r-RNA and proteins) as the machinery for protein production. Adapter molecules referred to as transfer RNA or t-RNA bring the individual amino acids to the site for protein synthesis. Typical rates of protein synthesis are twenty amino acids per second per ribosome. The quantity of RNA in a cell is controlled to match the cell's requirement for protein synthesis and can vary from about 5 to 20% of the cell's dry weight.

The other major fraction of the cell's mass is associated with the cell envelope. The purpose of the cell envelope is two fold. The first is to provide a selective barrier to allow the passage of selected molecules in or out of the cell's cytoplasm while preventing the entrance or escape of other molecules. The second is to provide structural integrity and cell shape. The latter is particularly important since the osmotic pressure inside a bacterial cell can be quite high.

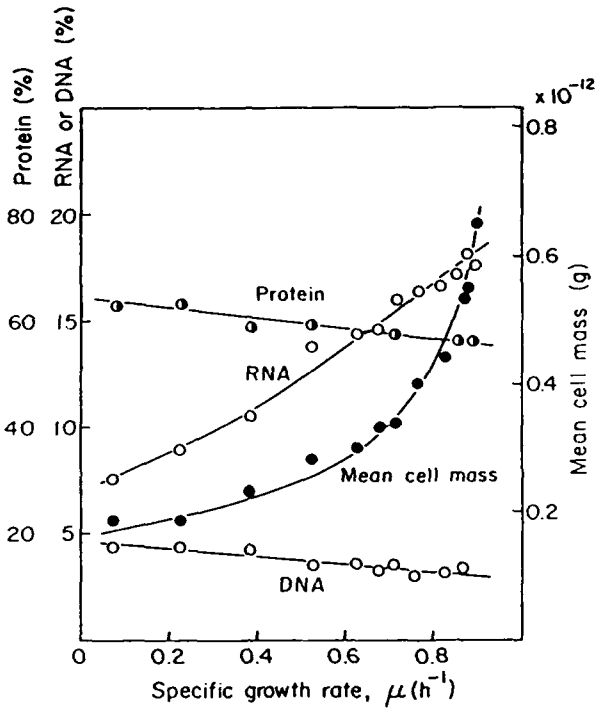


FIGURE 1 Effect of different rates of growth on the cell mass and chemical composition of *Aerobacter aerogenes* maintained in continuous culture (from Herbert, 1961; with permission).

The cell adjusts its composition to optimize either its chances of growth or of survival depending on the external environment. For example Figure 1 shows how bacterial composition can change with growth rate in a chemostat (or CFSTR).

These changes reflect changes in the biosynthetic capabilities or profile of enzymes. Enzyme activity which is no longer needed is modulated by small molecular weight effectors. For example feedback inhibition (usually the end product of the reaction pathway combines with the enzyme controlling entry into the pathway and alters the three dimensional shape of the enzyme and hence its activity) is a short term strategy. The longer term strategy is the regulation of enzyme synthesis often in the form of feedback repression. These mechanisms allow the cell to change its chemical structure in response to alterations in its external environment.

The key point from this section is that a living cell is a "catalyst" of high flexibility. The cell's biosynthetic capabilities can change significantly. A model that aspires to mimic a population's biosynthetic capability must be cognizant that the cells chemical structure is highly dynamic.

MODEL CHARACTERISTICS

A conceptual framework for classifying models of microbial populations was first suggested by Tsuchiya, Fredrickson, and Aris (1966). This framework has

been retained although there is no universal agreement on terminology. Harder and Roels (1982) have written a recent review on structured models which analyzes some of these concepts in detail.

Models can be deterministic or probabilistic. A deterministic model allows the exact prediction of future behavior based on specifying the current state vector. Generally a total population greater than 10,000 is sufficient to allow treatment of the system as deterministic. Since most systems of engineering interest contain microbial populations well in excess of 10,000 we will restrict our attention to deterministic models.

Models are generally "structured" or "unstructured". An unstructured model assumes that a single variable is adequate to describe the population. Typically this single variable is related to the *quantity* of biomass. Implicit in such models is the idea that the biosynthetic capabilities of the population are invariant.

A structured model divides the population into subcomponents. With a pure culture (only one species present) the addition of structure is most often accomplished by dividing the biosphere into two or more recognizable chemical subcomponents (e.g. RNA, protein, storage compounds, etc.). To more exactly characterize models, we introduce the term "chemically structured" to describe such a model. A mixed population (e.g. one containing two or more biological species) can be structured either chemically or by explicit recognition of each species. In the latter case the biosphere has structure in that subcomponents are recognized and the biosynthetic capability of the population will shift as the ratios of species changes. However, the model for each species need not be (and usually is not) chemically structured. We introduce the term "non-chemically structured" to describe such systems.

A special form of non-chemically structured models for pure populations is one that recognizes that a population is made up of distinct individuals. This form of structure is referred to as "segregation". A segregated model explicitly recognizes the distribution of properties among a population. A "non-segregated" model (also referred to as "distributive" or "continuum") views the cell mass as a lumped biomass which interacts as a whole with its environment. The non-segregated model is satisfactory as long as the properties of the culture can be adequately represented by averages. However, in some important and practical situations moments higher than first-order are important.

Consider a population of genetically-modified cells in which the number of plasmids is distributed non-uniformly. A plasmid is an extrachromosomal piece of DNA which can be manipulated to code for a foreign protein and then inserted into a cell. With an inducible promoter the amount of protein made per gene could be different in different cells since gene expression is a function of intracellular concentrations of inducer (essentially the same in all cells), of repressor (potentially variable if the repressor gene is encoded on the plasmid) and of binding sites on the gene of interest (related to plasmid number). Thus total protein production could be different in a population with a wide distribution of plasmid number per cell than in a population with a uniform population. An even more extreme example is the case where some cells can completely shed a plasmid. Some cells might have thirty or forty copies of a plasmid while some

have none. The cells without plasmids are freed of the "metabolic burden" imposed by spending resources on plasmid-encoded functions which do not benefit the host cell. Thus cells without plasmid can grow more rapidly than those with plasmids and can in a few generations displace the slower growing but productive plasmid carrying cells. Thus a culture where every cell has ten plasmids will be potentially much more productive than one where half the cells have twenty plasmids while the other half have none. The productivity of such cultures cannot be adequately forecasted by non-segregated cultures.

With these concepts of structure ("chemical structure", "non-chemical structured", and segregation) the role these models are required to fulfill can be delineated. First the concept of "balanced growth" needs to be defined. Campbell (1957) originally wrote "... it will be convenient to say growth is balanced over a time interval if, during that interval, every extensive property of the growing system increases by the same factor". Basically in balanced growth the composition of a "typical" cell is time invariant. Typical balanced growth situations are exponential growth (growth under nutrient saturated conditions) in batch culture and steady-state conditions in a CFSTR (or chemostat).

Intuitively one might suspect that a non-structured model might be satisfactory to describe a balanced growth situation since the biosynthetic capability of the cellular population is constant and only the quantity of biomass is required. On the other hand, most growth situations are associated with constantly changing biosynthetic capacities: transient responses in a single-state CFSTR, all behavior (steady-state or transient) in a multiple-staged CFSTR, and batch growth other than the exponential phase (see Barford and co-workers (1982)). Under such dynamic conditions one would suspect that a model would have to contain some type of structure to capture the behavior of a culture with variable biosynthetic capabilities. Such intuitive arguments are supported by mathematical derivations (originally demonstrated by Fredrickson *et al.* (1971) and more recently discussed by Harder and Roels (1982)). The mathematical derivations have shown that only structured models can possibly predict the growth of microbial populations during unbalanced growth. Unbalanced growth must be considered in designing bioreactors.

The design of a bioreactor system (number of reactors, batch or continuous, steady-state operation versus cyclic operation) and of a process control strategy for such systems is significantly aided by good models and clearly these *must be structured models.* Additionally such models offer important vehicles to test hypotheses about cellular control mechanisms and the prediction of the dynamic response of a bioreactor is a much more severe test of a model and its input mechanisms than is the prediction of only steady-state behavior.

This author believes that models of cell populations can only be complete if they contain both "chemical structure" and "segregation". Models which contain only "chemical structure", or "non-chemical structure" or "segregation" are significant improvements over unstructured models but in each case ignore potentially crucial features in the performance of the cell population. However, models which contain both "chemical structure" and "segregation" are complex and consequently forms with only one-type of structure are more easily constructed

and require less computational time. The engineer must strike the balance between reality and practicality.

CONSTRUCTION OF CHEMICALLY STRUCTURED MODELS

The first step to the construction of any mathematical model is the development of a conceptual model. Here an understanding of cell biology is indispensible.

The first decision is whether segregation is to be considered; for the purposes of this initial illustration assume that a non-segregated model is to be considered. Further, consider a batch reactor with an initially specified volume of fluid. The fluid is divided into two compartments—one abiotic and one biotic, representing the biomass. Clearly the growth of the cell population will result in an increase in the volume of the biotic mass, while the volume of the abiotic compartment will decrease. Many structured models have not adequately accounted for the changes in volume of these compartments.

The second decision is how to subdivide the cell biomass into compartments. For a chemically structured model at least two components must be specified. The other extreme would be a model which divided the cell into all of its chemical components. Since even a simple bacterium will have over a 1000 chemical species, such a model is unrealistic. The actual number chosen will represent a compromise between physical reality and mathematical complexity and will vary accordingly to the goals of the modeler. It is important that all chemical species be included by lumping them into a recognizable model component (e.g. all protein is represented by component P). Models intent on testing cellular control mechanisms will usually require the greatest detail.

The *third step* is to postulate the relationships among components. Typically a series of pseudo-chemical reactions are written. For example:

$$C_i + \alpha C_j + \cdots \longrightarrow C_k + \cdots \tag{1}$$

where C_i, C_j, and C_k are the intracellular or *intrinsic* concentrations of components i, j, and k, and α is a stoichiometeric coefficient. Equation (1) suggest that components i and j combine (ratio of amounts of j to i is α) to form k. Other unaccounted components may be involved in the reaction as products. The reaction represented by Eq. (1) is not a single enzymatic step but may represent many such steps. The pseudo-chemical reactions are used basically to keep track of the stoichimetry.

Since non-elementary reactions are used, the reaction kinetics are normally more complex than suggested by a cursory glance at Eq. (1). The kinetic expressions which are postulated must relate to the chemical conversion in Eq. (1) but will consider cellular control mechanisms and other interactions. For example a possible kinetic expression for Eq. (1) might be:

$$r_{ak} = k_a\left(\frac{C_i}{K_i + C_i}\right)\left(\frac{C_j}{K_j + C_j}\right) \tag{2}$$

where r_{ak} is the rate of formation of k by the ath process, k_a is rate constant

(units of gm k/L cell volume-h), C_i and C_j are the intrinsic concentration of i and j, and K_i and K_j are saturation parameters. Saturation kinetics were used in Eq. (2) to suggest enzyme kinetics and would give the appropriate behavior in the extremes. That is zero reaction if either i or j were absent and a constant rate of reaction when i and j are in excess. For a single enzyme with two substrates the reaction form in Eq. (2) can be derived as a special case of two-substrate enzyme kinetics (Mahler and Cordes, 1966). However, under most circumstances the conversion of i and j to k requires multiple enzyme steps. In such cases the approach of Kacser and Burns (1968, 1973) can be used to more rigorously derive an overall kinetic expression, at least for simple pathways. The expression in Eq. (2) may well be incomplete. For example, many enzymes or pathways are subject to feedback inhibition by the product of the pathway. In such cases Eq. (2) might be written as:

$$r_{ak} = k_a\left(\frac{K_p}{K_p + C_k}\right)\left(\frac{C_i}{K_i + C_i}\right)\left(\frac{C_j}{K_j + C_j}\right) \tag{3}$$

where C_k is the intrinsic concentration of k and K_p is an inhibition constant. Such a modification will mimic feedback inhibition by giving a decreased reaction rate in the presence of a high concentration of k.

Since model components i, j, and k may be aggregates of many chemical species, it is often impossible to exactly capture the basic chemical kinetics. The modeler must then be content with expressions that are consistent with observed overall mass balances (e.g. at the cell's maximum growth rate z gm of k are formed per h) and give the right general dependencies on all substrates and modifiers. If a single enzyme controls then it is possible to relate the saturation parameters in expressions like Eq. (2) to actual experimental values. Otherwise the saturation parameters should be related to the normal intracellular levels of components (i, j, and k).

Once the reaction kinetics have been determined an overall model can be formulated to predict the dynamic response of the culture. In writing the reactor balances the modeler must remain aware of the expanding nature of the biomass and the necessity to write the kinetics in terms of intrinsic concentrations. Clearly the enzymes within a cell only respond to intracellular concentrations. One exception is that the extrinsic concentration (gm/L or reactor volume) of substrates can be used. The use of the abiotic concentrations with substrates implies that the intracellular and extracellular concentration of substrate are in dynamic equilibrium. During transient response this assumption may be weak and explicit recognition of extracellular and intracellular concentrations of substrate must be made.

Although these constraints are fairly obvious, they have been ignored in many chemically structured models. Fredrickson (1976) was the first to point out this error. The correct formulation of a chemically structured, non-segregated model for a non-flow reactor is:

$$d/dt(m\hat{V}C_i) = m\hat{V}\sum_i r_{ai} \tag{4}$$

where: m = total biomass in the system at time t
$\hat{V}$ = volume of biomaterial per unit of biomass
C_j = mass of the jth component of biomaterial per unit volume biomaterial at time t—essentially the intrinsic concentration of j
r_{aj} = rate per unit volume of biomaterial at which the jth component of biomaterial appears (or disappears) because of the ath process

Normally $\hat{V}$ will be a constant and is essentially the reciprocal of the cellular dry weight density. Equation (4) can be rearranged to yield:

$$dC_j/dt = \sum_i r_{aj} - \mu C_j \tag{5}$$

where:

$$\mu \equiv 1/m\ dm/dt \tag{6}$$

The symbol μ is the specific growth rate, h^{-1}, and the term "$-\mu C_j$" represents the dilution of intracellular components brought about by growth. It is this term which has been neglected in many structured models.

The above equations could also be written in terms of mass concentrations. That is X_j (mass of jth component per unit mass of biomaterial) instead of C_j. Clearly $X_j = C_j\hat{V}$. Other formulations of the same concept can be used.

Although the above approach was illustrated for non-segregated models, the general concepts apply to at least some forms of segregated models (e.g. those involving construction of population models from single-cell models using a finite-representation technique).

Some characteristics of good chemically structured models are:

1. A minimum of adjustable parameters; most parameters should be determined directly from independent experiments or estimated by an objective series of rules,
2. Mathematically tractable,
3. High-fidelity to biological processes, and
4. Be experimentally verifiable.

With these basic concepts in hand we can turn our attention to examples of chemically structured models. It is not the purpose of this paper to exhaustively review all models but to illustrate various approaches to modeling.

EXAMPLES OF CHEMICALLY STRUCTURED MODELS

Two of the first chemically structured models proposed were those by Ramkrishna *et al.* (1967) and Williams (1967). Both were two component models. Ramkrishna *et al.* (1967) divided the cell population into a G-mass (RNA and DNA) and a D-mass (proteins (and by implication rest of the cell mass)). Williams (1967) split the cell mass into a synthetic component (primarily RNA) and a structural-genetic component (primarily DNA and protein).

These models were major conceptual breakthroughs in the art of modeling the growth of cell populations. Both models and many that followed have been flawed by not including intrinsic concentrations of internal cell components.

To illustrate these approaches consider Williams' (1967) original model. In the original model the following components were defined:

A = concentration (mass per unit volume of reactor) of limiting nutrient
D = concentration (mass per unit volume of reactor) of structural and genetic component
R = concentration (mass per unit volume of reactor) of structural and genetic
M = total biomass = $R + D$

All of the above are extrinsic concentrations. Williams (1967) assumed that component A was extracted from the medium and used to produce component R. Component D was assumed to be formed from R. The following equations were then postulated for R and D formation:

$$dR/dt = k_1AM - k_2RD \tag{7}$$

$$dD/dt = k_2RD \tag{8}$$

The above equations are conceptually incorrect since they are based on extrinsic concentrations of intracellular components.

The correct expression for Eq. (7) can be written:

$$\underset{\substack{\text{Rate of}\\ \text{change of } R\\ \text{concentration}}}{dR/dt} = \underset{\substack{\text{Biomass}\\ \text{concentration}}}{M} \quad \underbrace{\left[k_1A(M/M) - k_2(R/M)(D/M)\right]}_{\substack{\text{rate of}\\ \text{nutrient}\\ \text{uptake per}\\ \text{unit mass}\\ \text{of biomaterial}}} \quad \underbrace{}_{\substack{\text{rate of conversion}\\ \text{of } R \text{ into } D \text{ per}\\ \text{unit biomaterial}}} \tag{9}$$

or

$$dR/dt = k_1AM - k_2\frac{RD}{M} \tag{10}$$

which is clearly different than Eq. (7). Equation (4) could also have been used directly to derive Eq. (10): i.e.

$$d(R/M)/dt = k_1A(R/M + D/M) - k_2\frac{RD}{MM} - \mu R/M \tag{11}$$

Equation (11) reduces directly to (10) when the differentiation on the left hand side of (11) is carried out, and the definition for μ ($\mu \equiv 1/M\, dM/dt$) is substituted into (11).

The correct expression for the D component is

$$dD/dt = M\left[k_2R/M \cdot D/M\right] = k_2RD/M \tag{12}$$

which is different from Eq. (8).

The above discussion says nothing about the reasonableness of the kinetic expressions chosen by Williams (1967), but only the correct conversion of those kinetic concepts into mathematical terms. A potential point of disagreement, for example, could be whether biomolecular kinetics or saturation kinetics would be preferable.

A better example of a simple chemically-structured, non-segregated model is

the one described by Harder and Roels (1982). They proposed a three-compartment model. Three-compartment models have one distinct advantage over two-compartment models; the addition of an extra component makes it much easier to relate model components to chemical species and to the organization of cell structure. Consequently the estimation of parameters from experimental data is far easier with three compartment models than two.

The structure of their model is depicted in Figure 2. The K compartment is RNA, the G compartment is protein, and R is the remainder of the biomass consisting primarily of carbohydrates (including storage compounds), lipids, and precursors. It is presumed that the limiting substrate (energy and carbon source) enters the cell population and is immediately incorporated into the R component. The K and G components are made from precursors in the R component, and K and G can be degraded ("turn-over") to yield precursors which return to the R compartment.

The kinetic expressions suggested by Harder and Roels (1982) were based explicitly on assuming a pseudo-steady-state for ATP (the cell's energy currency) and precursors. The justification for such a hypothesis is that the relaxation times for adaptation of ATP and precursor concentrations are much less than for the other components (K, G, and R). Yield coefficients (essentially stoichiometric coefficients) based on ATP and precursors can be estimated from a knowledge of basic cell biochemistry and coupled to the requirements for the uptake of substrate.

This model satisfies most of the constraints of a good model and has the virtue of being computationally simple. One disadvantage of the model includes the assumption of a very tight coupling of ATP production with ATP needs—an assumption which ignores energy-spilling reactions (see Stouthamer, 1979).

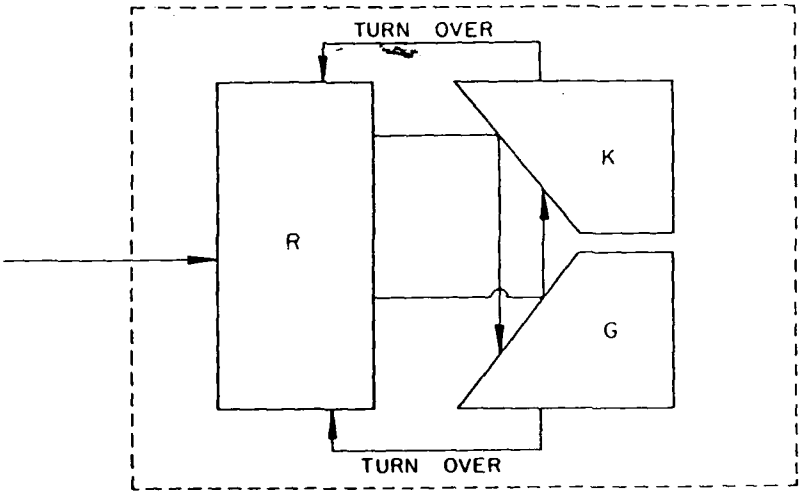


FIGURE 2 Schematic representation of a three-compartment chemically structured growth model (from Harder and Roels, 1982; reproduced with permission).

Further, the kinetic relationships among K, G, and R are based on relationships from steady-state continuous culture experiments, and consequently may be inadequate for predicting transient responses. The other criticism is that the model was to be applied to an activated sludge system; consequently, the authors implicitly assumed that parameters derived primarily from studies on *E. coli* in pure culture could represent the average parameters from a very diverse group of organisms. Interactions among species in the population were ignored. Some, such as predation by protozoa on bacteria, would be expected to alter parameter values and would give the observed value of the parameters a time dependence since the ratio of protozoa to bacteria can be time dependent. Nonetheless such three compartment models represent an attractive approach, particularly for pure cultures, when computational speed is important and the available data base for parameter estimation is small.

Another approach to structure is to view the cell as a process and provide structure by dividing the process into distinct steps. Chiam and Harris (1982) have developed a simple model of this form (see Figure 3). This model was tested by comparison to the data of Mor and Fiechter (1968) (see Figure 4) and the model predictions are in reasonable accord. Unlike the simple unstructured Monod equation, the model predicts a decrease in yield coefficient at both high and low dilution rates as observed in this phenol utilizing system. This comparison is not a

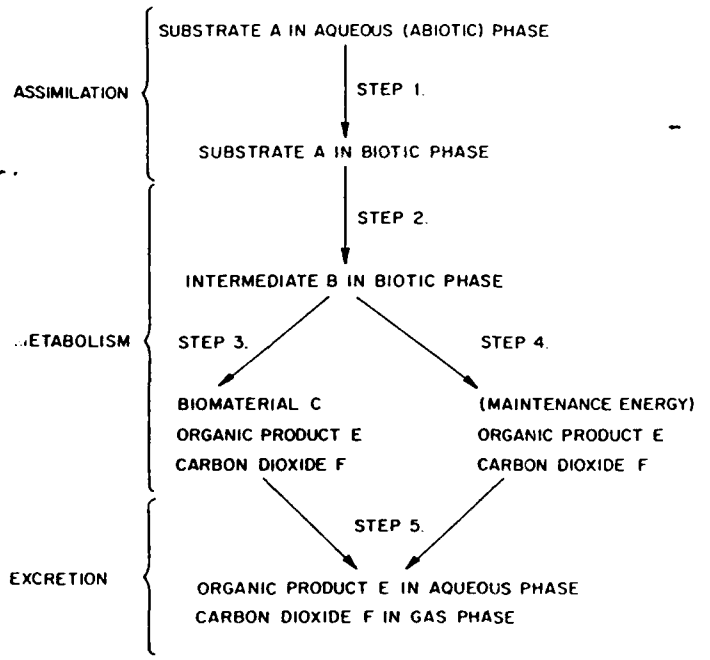


FIGURE 3 Schematic representation of a model structured from a process perspective (from Chiam and Harris, 1982; with reproduced permission).

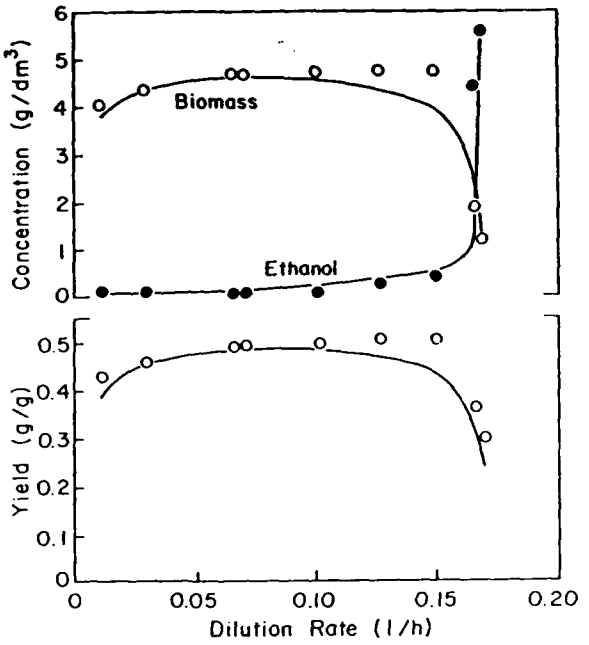


FIGURE 4 Comparison of predictions from Chiam and Harris (1982) with the data of Mor and Fiechter (1968). Points refer to experimental data and the continuous line to model predictions. The parameter values were obtained from a fitting technique (from Chiam and Harris, 1982; reproduced with permission).

very stringent test of the model's capabilities; the prediction of steady-state behavior is far easier to achieve than the prediction of transient responses.

A novel approach to structure which avoids the use of complex kinetic pathways has been advocated by Ramkrishna (1983). This "cybernetic perspective" assumes that the cell behaves optimally with regard to the allocation of existing resources among parallel enzyme-synthesis systems. This approach has been applied with some success to growth in multiple substrate systems.

A large number of other structured non-segregated models have been written (see Bazin, 1982 for other examples). Such models may be fitted to transient response data and give reasonable correlations of experimental data. However, such models cannot give *a priori* quantitative predictions of population response to transient conditions over a range of growth conditions using inoculum from different environments (see Daigger and Grady, 1982). The approach described in the next section, single-cell models as a basis for population models, can make such predictions.

SINGLE-CELL MODELS AS A BASIS FOR POPULATION MODELS

The previous models have looked at the cell population as a lumped biomass. An alternative view is to recognize that a cell population represents a complex

averaging of the biosynthetic capabilities of individual members of the population. The remainder of this paper is devoted to the single cell perspective. A type of chemically-structured, non-segregated model is created when we construct a single-cell model and view that single-cell as representing an "average" cell in the population. A population model containing both chemical-structure and segregation can be constructed by using a finite-representation technique whereby a fraction of the total population is represented by a single-cell model. Such models can predict the distribution of biosynthetic or other properties within the population. Shuler, Leung, and Dick (1979) first suggested this possibility as a means to avoid the mathematical difficulties of solving the integro-differential equations involved in writing a chemically-structured segregated model from the population point of view.

The use of single-cell models to predict the average behavior of a cell in a population has several potential advantages. Certainly such models make it easier to incorporate biochemical structure into a model; cell geometry (e.g. surface to volume ratios) can be readily incorporated; and temporal events or spatial effects can be included that would be virtually impossible to incorporate into lumped biomass models. As discussed, the single-cell modeling approach can be readily extended to develop chemically-structured segregated models of cell populations.

One of the first examples of single-cell models was that suggested by Von Bertalanffy (see Tsuchiya *et al.*, 1966) in 1942. In his model growth was a result of competition between the processes for nutrient assimilation and for endogenous metabolism. Many examples of other single-cell models could be given (see Shuler and Domach, 1983), but most are flawed by unrealistic constraints on cellular growth mechanisms or do not reflect current understanding of cell biology.

An ambitious type of single-cell model which is biologically sound is one proposed by Lee and Bailey (1984a) for the production of plasmid encoded proteins. A schematic of the model is given as Figure 5. The biochemical events controlling the replication of the plasmid $\lambda\, dV$ are well known and can be accurately modeled (Lee and Bailey, 1984b, 1984c). In addition accurate models have been formulated for the lac promoter (Van Dedem and Moo Young, 1973; Imanaka and Aiba, 1977; Gondo *et al.*, 1978). An inducible promoter is a chemical switch which can turn on or off the transcription and translation of information coded on a gene into a functional protein. Lee and Bailey (1984a) have incorporated detailed molecular models of plasmid replication and lac promoter with semi-empirical expressions for host cell functions. The model assumes regular plasmid segregation (i.e. each daughter cell formed by binary fission receives an equal number of plasmids). This model is a chemically structured non-segregated model, since it is based on a single-cell as an average cell in the population representative of the total population.

The model by Lee and Bailey (1984a) has a larger number of parameters than those models previously discussed. However, these parameters can largely be evaluated from independent experiments, and the structure of the equations for plasmid replication and for gene expression are consistent with our knowledge of molecular biology. The model is a quite powerful one. Models with a large number of *independently determinable* parameters embedded in a kinetic structure

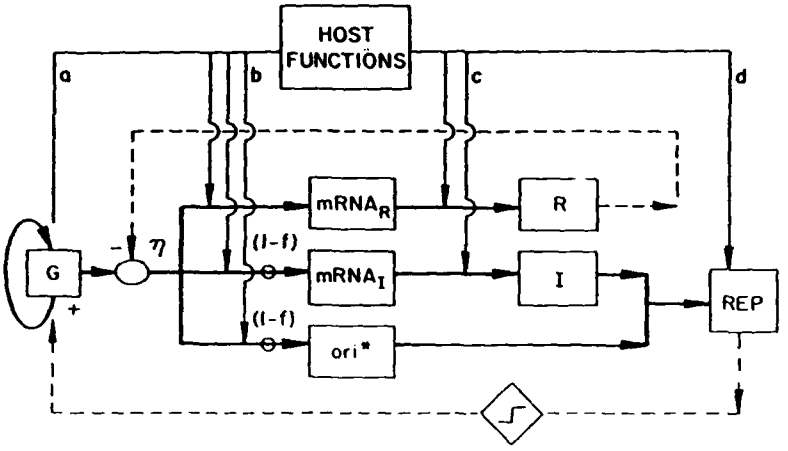


FIGURE 5 Schematic of Lee and Bailey's (1984a) model for λdV plasmid replicon and its interaction with *E. coli* host cell functions. Definition of symbols are: I = initiator protein, R = repressor protein, G = number of plasmid molecules per cell, ori* = activated origin, REP = replication complex, $mRNA_R$ = messenger RNA for repressor protein, $mRNA_I$ = messenger RNA for initiator protein, F = termination efficiency, and η = transcription efficiency. The dotted line shows the regulatory action of repressor protein and initiator protein. Traces (a), (b), and (c) represent replication, transcription, and translation processes, respectively. Line (d) indicates the involvement of replication protein of the host cell in formation of replication initiation complexes (from Lee and Bailey, 1984a; with permission).

consistent with experimental observations are, at least in principle, intrinsically superior to models with three or four empirical parameters, particularly when the empirical model structure cannot reflect known kinetic interactions. This superiority is reflected primarily in testing hypotheses concerning cellular mechanisms and in making predictions on growth dynamics different from the initial data base. The construction of models with complex kinetic structures are generally not practical for engineering applications unless a fairly extensive data base already exists. Fortunately the explosion in the literature in cellular and molecular biology provides an ample and generally untapped base for such model building. Once the engineer is readily acquainted with such literature, the construction of complex models is not difficult or unduly time-consuming.

Adequate experimental data to rigorously test the model predictions under a wide variety of growth conditions are not currently available. However, Lee and Bailey (1984a) have presented a comparison (see Figure 6) of predictions of recombinant protein activity to data on the activity of plasmid encoded β-lactamase. The model predictions are generalized for the production of an unspecified protein, so that a comparison to β-lactamase requires an estimate of the specific rate of decay for β-lactamase. Satisfactory agreement of predictions and experiment were obtained for a decay rate of *ca.* 0.001 min^{-1}.

However, the model can be criticized, primarily for the simplifying assumptions involved in the integration of the kinetic expressions for plasmid replication and gene expression with host cell functions. The host cell functions are empirical

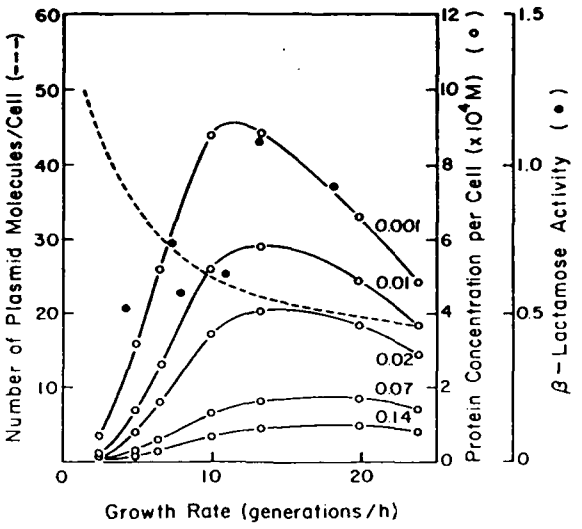


FIGURE 6 The effect of growth rate on cloned-gene protein. Assumed decay constants of product protein are given near each curve (units are min^{-1}). The experimental data (●) were taken from Engberg and Nordstrom (1975). (Figure from Lee and Bailey, 1984; reproduced with permission.)

relationships for cell replication, transcription, and translation processes. They are soundly based on experimental observations. These empirical expressions require as input a value for the cell's growth rate, μ. The host cell functions do not respond explicitly to changes in the abiotic environment such as concentration of a limiting-nutrient. Thus the model is restricted to the range of growth rate used to establish the empirical expressions and may well fail under transient growth conditions where substrate concentrations and growth-rate become decoupled.

Another potential weakness is that the model assumes that the host cell functions are not altered by interactions with the replication of plasmids and expression of *r*-protein synthesis. The presence of the "metabolic burden" posed by plasmid encoded functions will alter cell growth rates. Although $\mu = 0.8$ gen/hr in a glycerol minimal medium for the wild-type cell, the growth rate of the genetically-modified cell will almost certainly be less. How much less is difficult to predict, and it is not clear to this author how the model could predict such differences or how to use the model without an experimental knowledge of the actual growth rate of the recombinant organism.

These comments suggests ways in which the model is limited. The reader should nonetheless be aware of the pioneering nature of the model and of its importance. This model is the best one currently available to explore important questions on mechanistic control of plasmid copy number and of gene expression when the growth rate is altered. Careful attention to this type of model is warranted by both experimentalists and modelers.

The model with the greatest fidelity to cellular biochemistry is probably the one

suggested by the Shuler and colleagues (Shuler, Leung, and Dick, 1979; Shuler and Domach, 1983; Domach *et al.*, 1984; Domach and Shuler, 1984a, Lee, Ataai, and Shuler, 1984; Ataai and Shuler, 1984a) for the bacterium, *Escherichia coli*. A schematic of the single-cell model is given in Figure 7 as well as a definition of the model components.

In this approach the cell is treated as an expanding reactor free to change shape and volume and to respond *explicitly* to changes in glucose or ammonium concentrations in a minimal medium. Glucose, and/or ammonium ion, can be the limiting nutrient(s). This model contains sufficient detail that it provides a good tool for quantitatively testing the plausibility of cellular control mechanisms. It provides a holistic view of the cell and questions about any subcomponent can be

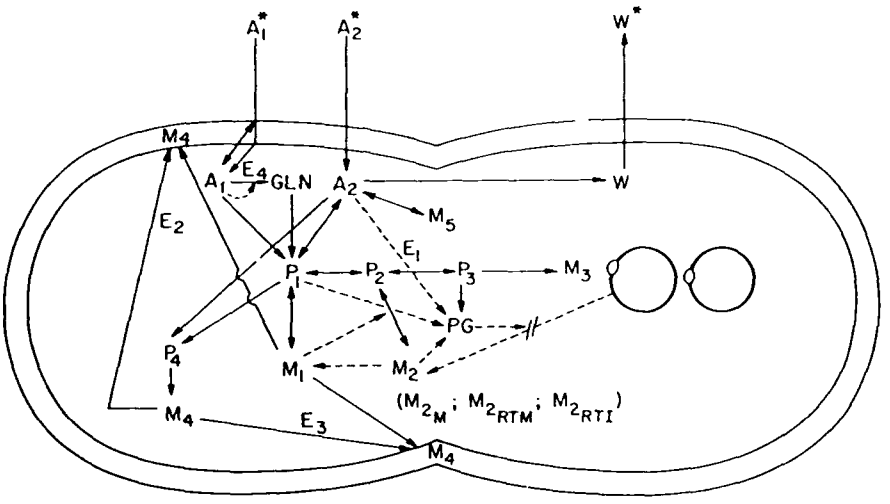


FIGURE 7 An idealized sketch of the model of *E. coli B/rA* growing in a glucose-ammonium salts medium with glucose or ammonia as the limiting nutrient. At the time shown the cell has just completed a round of DNA replication and initiated cross-wall formation and a new round of DNA replication. Solid lines indicate the flow of material, while dashed lines indicate flow of information. Reproduced with permission from Shuler and Domach, 1983.

A_1 = ammonium ion
A_2 = glucose (and associated compounds in the cell)
W = waste products (CO_2, H_2O, and acetate) formed from energy metabolism during aerobic growth
P_1 = amino acids
P_2 = ribonucleotides
P_3 = deoxyribonucleotides
P_4 = cell envelope precursors
M_1 = protein (both cytoplasmic and envelope)
$M_{2_{RTI}}$ = immature "stable" RNA
$M_{2_{RTM}}$ = mature "stable" RNA (*r*-RNA and *r*-RNA—assume 85% *r*-RNA throughout)
M_{2_M} = messenger RNA
M_3 = DNA
M_4 = non-protein part of cell envelope (assume 16.7% peptidoglycan, 47.6% lipid, and 35.7% polysaccharide)
M_5 = glycogen
PG = *ppGpp*
E_2, E_3 = molecules involved in directing cross-wall formation and cell envelope synthesis—the approach used in the prototype model was used here but more recent experimental support is available
GLN = glutamine
E_1 = glutamine synthetase
*—the material is present in the external environment.

examined in detail and still be related to the observable performance of the whole cell. The model also provides a tool to test the translation of *in vitro* results to *in vivo* ones.

The cell is treated as an *expanding reactor* and mass balances can be written for each component. The equations are too complicated to justify a detailed listing here but such listings are available in original references (Shuler and Domach, 1983; Domach *et al.*, 1984). As an example, consider the equation for DNA synthesis:

$$dM_3/dt = M_3\left(\frac{P_3/V}{K_{M,P_3} + P_3/V}\right)\left(\frac{A_2/V}{K_{M,A_2} + A_2/V}\right)F \tag{13}$$

where M_3 is *amount* (not concentration) of DNA, t is time, P_3 is the *amount* of deoxynucleotides, A_2 is the *amount* of intracellular glucose and low molecular organics, V is the cell volume, F is the number of replication forks, μ_3 is a rate constant for the maximum rate of DNA formation per fork, and K_{M,P_3} and K_{M,A_2} are saturation constants. The value of μ_3 is readily determined from published values for the size of the chromosome, the number of forks present, and measurements of the time required for the fork to transverse the chromosome under maximum growth conditions. A value for K_{M,P_3} is directly available from the literature. The value for K_{M,A_2} was estimated as 1/25th of the normal intracellular value of A_2 at maximum growth in minimal medium. The determination of F involves a separate set of equations associated with the control of initiation of chromosome replication. The model was used to test the plausibility of six hypotheses about mechanisms controlling initiation and found only one mechanism and its modifications could make quantitatively plausible predictions about the observable pattern of chromosome replication.

The model is complex with a large number of kinetic parameters. Most of the 88 parameters could be estimated independently as described in the previous paragraph for Eq. (13). Four parameters, associated with the rate of cross-wall formation were determined by running the model at conditions supporting maximum growth and at a glucose concentration supporting approximately half-maximal growth rate. The values of these parameters were manipulated primarily to obtain the right septation rate as evidenced by reasonable predictions on cell geometry and the length of time between chromosome termination and cell division and the actual time of chromosome replication. Extension of the model to predictions of the response to nitrogen-limitation required no adjustable parameters—only values obtained from *in vitro* experiments on the key enzymes for ammonium incorporation into the cell. Thus without the addition of any adjustable parameters the model can adequately predict system response to growth at a variety of ammonium concentrations (Shuler and Domach, 1983). The model makes quite reasonable predictions of steady-state cell behavior. See for example Figure 8.

The model has also demonstrated the capacity to make predictions about cellular behavior prior to the actual confirming experimental observations. In the prototype model (Shuler *et al.*, 1979) we found that the cell width varied slightly during the cell cycle. Marr, Harvey, and Tentini (1966) had measured cell widths

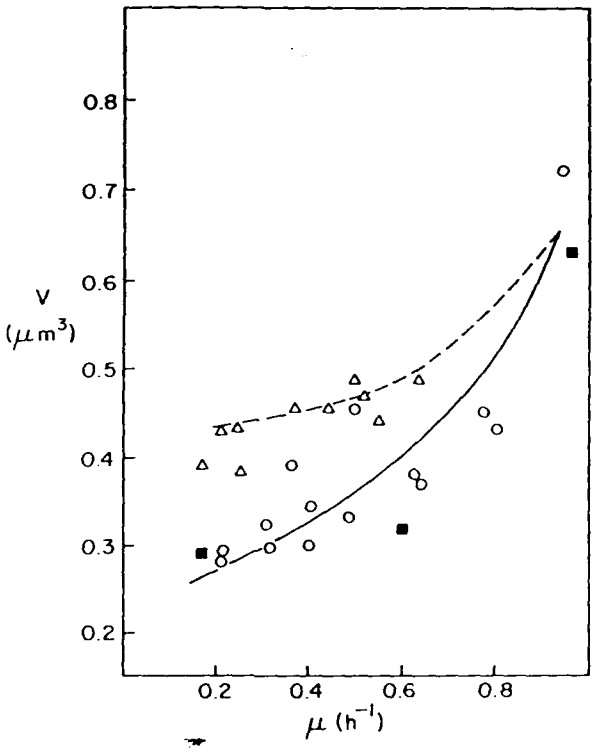


FIGURE 8 Variation of cell volume with steady-state growth rate. Model predictions for glucose-limited chemostat growth are indicated by the solid line while experimental data (Coulter Counter measurements) for glucose limited growth are indicated with a dashed line, while the corresponding data are given as Δ. Data from Helmstetter (1974) for glucose-limited growth is given by ■. Reproduced with permission from Shuler and Domach, 1983.

and concluded that the cell width was constant throughout the cell cycle. Their technique, however, was precise to only $\pm 0.03\ \mu m$ and the variation within the model cell cycle was nearly within this limit. The model's prediction of varying width during the cell cycle was confirmed by experimental evidence from Trueba and Woldringh (1980). They observed about a 8% difference in cell width between the extremes of the division cycle with somewhat larger percent deviations with slow growing cells. The minimum width was observed to occur just before cell constriction was readily visible. In Domach *et al.* (1984) the maximum predicted variation in width was 8, 9, and 9% at specific growth rates of 0.95, 0.51, and $0.24\ h^{-1}$, respectively. The minimum width was predicted to occur after the initiation of cross-wall synthesis but before 50% constriction with the actual timing of the minimum point being dependent on growth rate. The ability of the model to make significant quantitatively accurate predictions about cell behavior prior to the actual observation of such behavior provides strong support for the fidelity of the model to actual cell biology and for the robustness of the model.

The model has been extended to anaerobic growth situations by altering the mechanism for energy generation in the cell (Ataai and Shuler, 1985a). Carbon and electron balances must be added so that the amount of ATP and reducing power generated meet the requirements for consumption. The anaerobic model, which left the non-energy related part of the aerobic model intact, can make accurate predictions of cell size, cell composition, growth rate, and amount and composition of metabolic end products (primarily acetate and ethanol and succinate) (see Figure 9). One conclusion from this work is that all non-energy producing processes remain unaltered in kinetics when a faculative organism switches from aerobic to an anaerobic growth. Thus the base model has proven to be a robust one capable of extension to growth conditions significantly different from those used in its initial derivation.

A single-cell model when used to represent an average cell is a form of a chemically-structured non-segregated model. Certain behavior can only be predicted by recognizing the differences among individuals in a population. Unlike other non-segregated models a single-cell model can be easily incorporated into a population model (Shuler, Leung, and Dick (1979)). Three examples of the

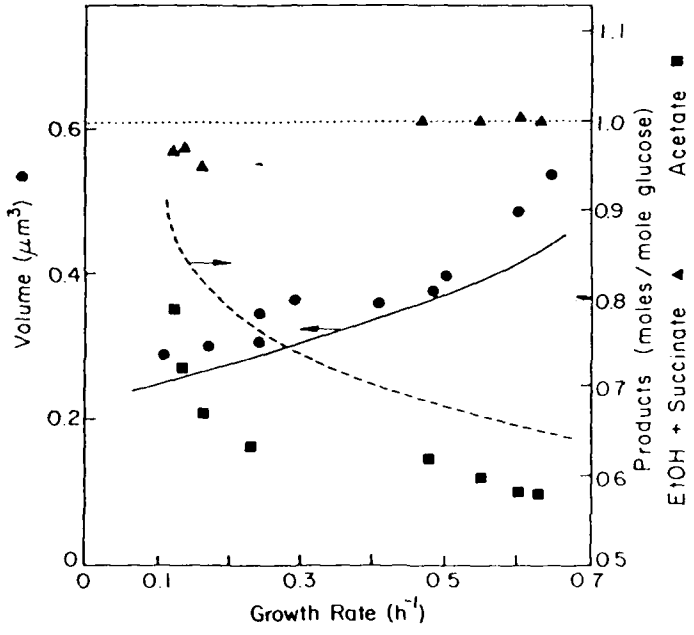


FIGURE 9 Effect of growth rate on cell volume and by product formation for anaerobic glucose-limited chemostat growth of *E. coli*. The basis of the experiment is 1 mole of glucose fermented. The dotted line is the model's prediction of the total ethanol plus succinate formed while ▲ represent experimental measurements. The dashed line is the model's prediction of acetate formed while ■ represents the amount of acetate experimentally measured. The predicted cell size is indicated by the solid line and the cell volume measured by the Coulter Counter is given as ●. No adjustable parameters were used in obtaining these results. Data from Ataai and Shuler, 1985a

construction of population models are given by Nishimura and Bailey (1980), Alberghina, Martegani, and Mariani (1982) and Domach and Shuler (1984b).

Nishimura and Bailey (1980; 1981) have demonstrated the feasibility of constructing a population model from a knowledge of a single-cell kinetics and mechanisms for control of replication. The general approach is detailed in Figure 10. Semi-empirical relationships for the timing of cell division, cell growth, and the timing of initiation of DNA replication were used as the basis for the single-cell model for *E. coli*. The single-cell model is chemically structured in that DNA content and total cell mass are both included. The DNA configuration at any time is known and consequently gene dosage can be readily calculated. Nishimura and Bailey (1980) accomplished the difficult task of determining the analytical solution of the distribution of cell mass and DNA content in a population of *E. coli*. The required input for the model was the cellular specific growth rate $\mu(t)$. The resulting model was able to predict the correct trends (a non-quantitative comparison) in transient responses to shift-up conditions (see Figure 11).

The chief potential disadvantage in the model is that the cellular specific growth rate $\mu(t)$ is considered to be a known function of time. Additionally the timing of the cell division cycle is deterministic once $\mu(t)$ is specified. In reality a probabilistic approach is necessary since there is a distribution of division times among cells in balanced growth for a culture at a known growth rate. Consequently the property distributions are likely to be broader than prediction by this model.

A population mode with a more detailed chemical structure and the ability to explicitly respond to external changes in nutrient levels has been constructed from our single-cell model (Shuler and Domach, 1983; Domach and Shuler, 1984b).

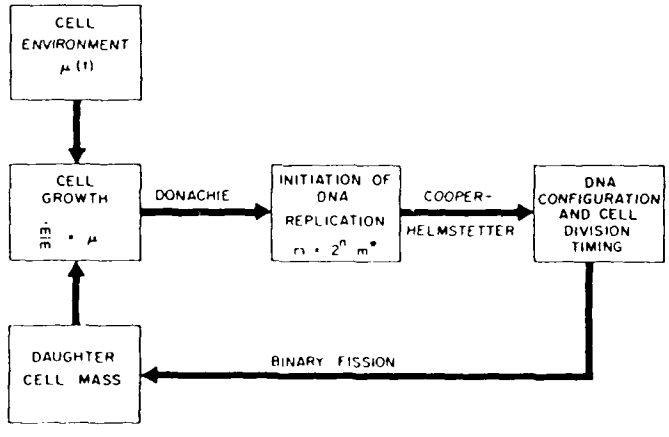


FIGURE 10 Schematic of a model for coordinated mass and DNA synthesis and cell division for individual cells of *E. coli*. Donachie refers to a simple model relating cell growth to initiation of chromosome replication (Donachie, 1968). Cooper-Helmstetter refers to a model relating growth rate to the timing and rate of replication of the chromosome and the timing of cell division (Cooper and Helmstetter, 1968) (reproduced with permission from Nishimura and Bailey, 1980).

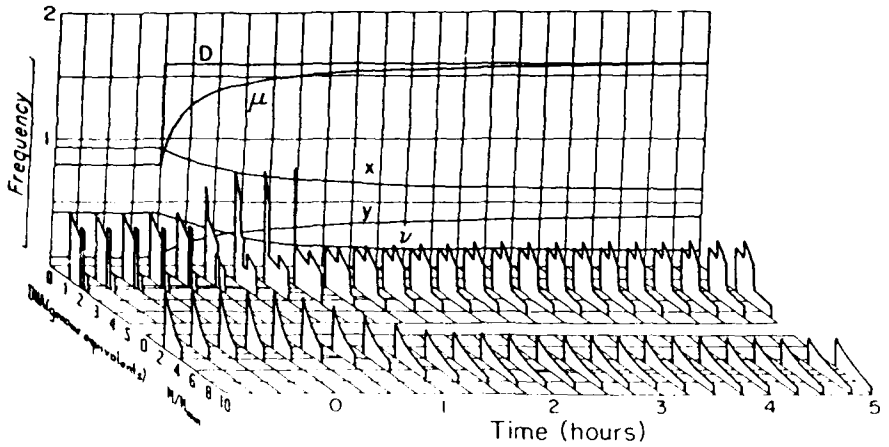


FIGURE 11 Calculated response of a bacterial CSTR to a step change in feed flow rate. The model is that of Nishimura and Bailey (1980) for *E. coli*. Curves D, μ, X, Y, and ν are dilution rate, specific growth rate, and dimensionless cell mass, substrate, and cell number concentrations, respectively. (Reproduced with permission from Nishimura and Bailey, 1981).

The first step in constructing such a model requires the postulation and confirmation of a cause for asynchrony in the cell division cycle. Domach and Shuler (1984a) postulated that the random variations in the quantity of enzyme responsible for cross-wall formation could result in the observed distributions in cell division cycle times (τ_d) and cell size at division (m_d). By including in the single-cell model a random number generator for the quantity of enzyme produced for cross-wall formation, the observed distributions in τ_d and m_d can be predicted (see Figure 12). The relative imprecisions predicted by the model (ratio

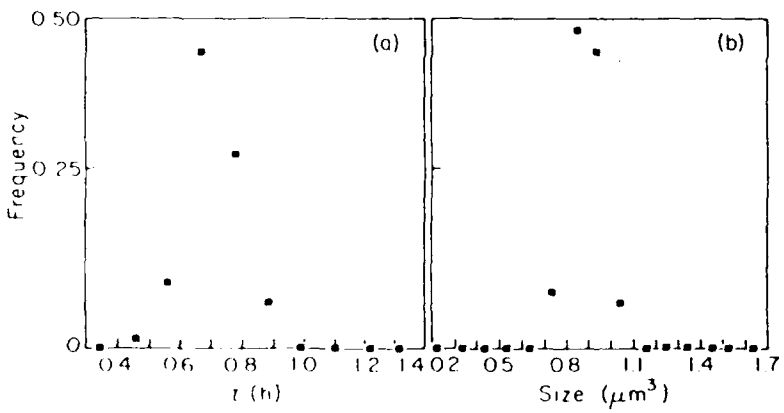


FIGURE 12 Prediction of variations of fission cell size and cycle time for individual cells of *E. coli*. The cycle time frequency (plot a) has a greater deviation about its mean than fission size (plot b) and the ratio of the coefficients of variation of cycle time and fission size is 1.7 which agrees well with reported observations (reproduced with permission from Domach and Shuler, 1984a).

of coefficient of variations of cycle time to fission size is 1.7) agree well with experimental observations (Schaechter *et al.*, 1962). The correlation coefficient calculated for parent-offspring cycle times was −0.4 which is slightly more positive than the theoretical value of −0.5 calculated by Koch and Schaechter (1962) but agrees well with the experimental value cited (−0.37). The negative correlation of parent-offspring cycle times has been widely observed.

Using a single-cell model (Shuler and Domach, 1983; Domach *et al.*, 1984; Domach and Shuler, 1984a; Ataai and Shuler, 1985a) with suitable modifications to account for asynchrony in the individual cell cycle allows for the construction of a population model. We use a finite-representation technique in which the behavior of a small fraction of the population is represented by a single computer cell. The key item for the reader to note is that *no adjustable* parameters are added to the single cell model to make predictions on the behavior of a population in a bioreactor.

Domach and Shuler (1984b) have used a population model consisting of 225 computer cells and mass balances on a CFSTR to predict the response of a culture to upsets in reactor conditions (e.g. feed flow rates or substrate concentrations). A flowsheet summarizing the operation of the population routine is given in Figure 13. Such a model accurately predicts the transient behavior of the culture (see

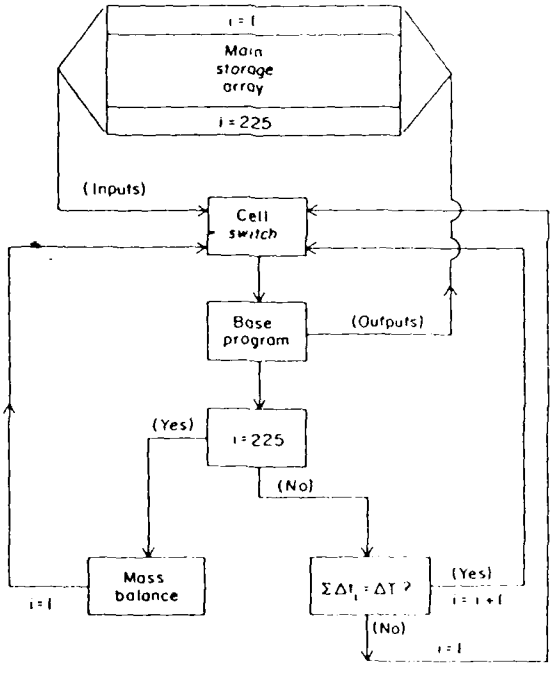


FIGURE 13 A flow sheet for the operation of a finite-representation population routine. Each computer cell ($i = 1$ to 225) is activated for a set "on time" ($\Delta T = 0.01$ h). Once all 225 cells have reacted with the environment, a mass balance is done to determine substrate levels and washout of cells. (Reproduced with permission from Domach and Shuler, 1984b).

Figures 14 and 15) with respect to overall performance (e.g. substrate consumption) and the time-dependent change of property distributions in the culture (e.g. size distributions). Similarly Ataai and Shuler (1985b) have subjected a single cell model for the anaerobic growth of *E. coli B/rA* to perturbations in flow rate and glucose concentration in a glucose-limited chemostat. A comparison of model predictions of RNA content and unconsumed glucose concentrations to actual data for a step change in feed glucose concentration (1.0 q/L to 1.88 q/L) is given in Figure 16. For a flow perturbation (dilution rate of 0.38 h^{-1} to 0.55 h^{-1}) the disturbance lasts longer (8 h vs 3.5 h), and the amount of unconsumed glucose reaches a higher level (ea. 180 mq/L vs 90 mg/L). The distribution of cell sizes is also well predicted in both cases.

To understand the importance of this approach to chemically structured-segregated models, the reader must realize that the attempt to generate models of similar characteristics from the population-balance point of view has been impossible—largely because the resulting equations have proved mathematically intractable.

Other attempts to develop models for the *a priori* quantitative predictions of transient behavior of cell populations have been generally less successful than the approach of using biologically detailed single-cell models in a finite-representation scheme. Daigger and Grady (1982) have recently determined that the ability of most other mathematical models (particularly structured models that presume that the cellular level of RNA is always growth rate limiting) to predict transient responses is inadequate. Daigger and Grady (1982) concluded that the RNA level

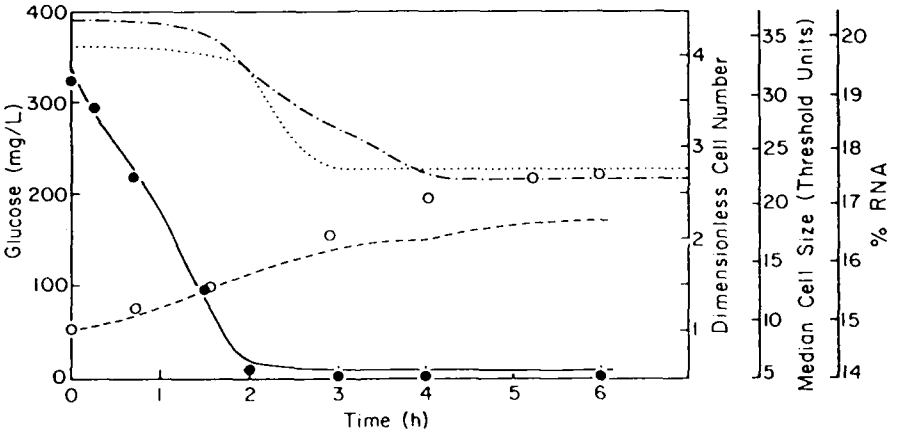


FIGURE 14 Transient response of aerobic glucose-limited chemostat with *E. coli* to a step decrease in dilution rate from 0.91 h^{-1} to 0.65 h^{-1} at time = 0. Model predictions of substrate concentration are given by the solid line, while experimental values are given by ●. Predictions of dimensionless cell number are given by the dashed line, while the open circles (O) denote measured values. Predicted variations in the RNA content of the cell are given by the dotted line. Measured values of the median cell volume are reported in terms of threshold units on the Coulter Counter and displayed as the dash-dot line. The initial lag in cell response and physiological state is predicted by the model (e.g. RNA content) and is consistent with the experimental measurements on cell volume. Data and predictions taken from Domach and Shuler (1984b).

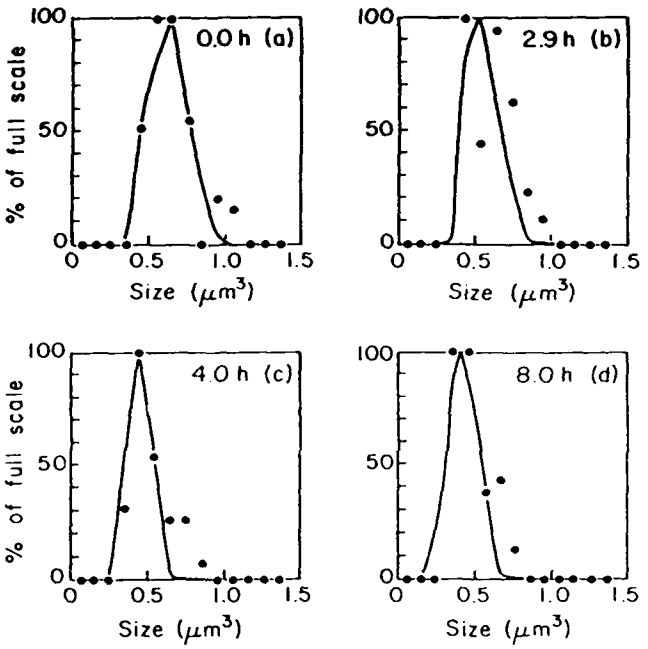


FIGURE 15 Shift in predicted and observed cell size distributions due to a step decrease in dilution rate of $0.91\,h^{-1}$ to $0.65\,h^{-1}$ in an aerobic glucose-limited chemostat with *E. coli*. A finite-representation technique using 225 single-cell models was used (Domach and Shuler, 1984b). The model predictions are given as a histogram and each ● represents the relative amount of the population in that cell class. The Coulter Counter measures cell volume and displays a continuous trace, and the experimentally measured distribution is indicated by the cell line. A smoother prediction of the cell size distribution can be obtained if the number of single-cells in the population routine is increased. The medium cell volumes determined were: 0.65 μm^3, 0.54 μm^3, 0.46 μm^3, and 0.42 μm^3 at $t = 0$, 2.88 h, 4.0 h, and 8.0 h, respectively. The model could be used to output distributions of any parameter (e.g. RNA, DNA, etc.). Reproduced with permission from Domach and Shuler, 1984b.

did not exclusively control the nature of the transient response and that other unidentified components varied with the steady-state specific growth rate that the culture had been subjected to prior to shift-up.

Their observations are consistent with the view that a cell contains many potentially growth-rate limiting steps and the actual controlling steps depend on cell history and the nature of the shift. This view supports our contention that the accurate *a priori* prediction of transient response requires models which mimic the actual control systems in the cell; all of the major control systems must be included because of their interactions and the potential for switches in the combination of systems which may be growth-rate controlling. Thus a model which aspires to accurate dynamic predictions over a range of growth conditions must be complex. We believe that our single-cell model may be near the *minimum level of complexity* to allow *a priori* quantitative predictions of transient behavior over a reasonable range of experimental conditions. We currently have

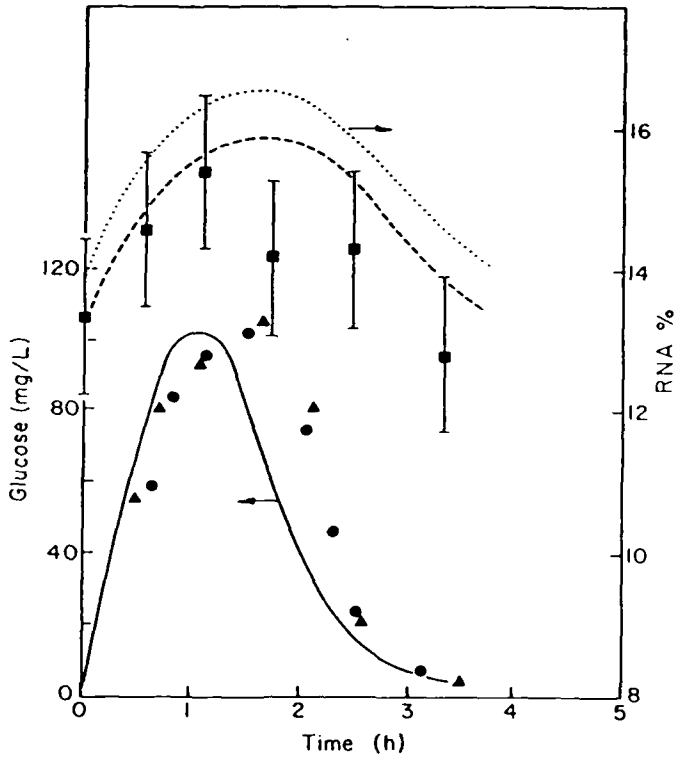


FIGURE 16. Transient response of an anaerobic glucose-limited chemostat with *E. coli* to a step change in glucose feed concentration (from 1.0 g/*L* to 1.88 g/*L*). The predicted extrinsic concentration of glucose in the reactor is shown by the solid line; the experimental values from two independent but identical experiments are given as ● and ▲. The predicted transient in cellular RNA content is given as a dotted line; the experimental values as ■. The dashed line indicates the model prediction if those predictions are normalized to the same initial RNA value as determined experimentally. It is important to note that the model predictions are made without using adjustable parameters or a fitting procedure. (Data from Ataai and Shuler, 1985b).

initiated a sensitivity analysis to better discern what steps may be most strongly growth-rate affecting under a variety of growth conditions. Such an analysis may suggest opportunities for reducing the complexity of our model, but we anticipate that the chemical structure can*not* be greatly altered without adversely effecting the quality of predictions for transient behavior.

The use of finite-representation techniques offers other advantages. For example such a model generates by itself the underlying distributions of cycle time and division mass found in real populations. The partial integro-conservation equations in age- and mass-structured models (e.g. Fredrickson and Tsuchiya, 1963; Eakman *et al.*, 1966) require *a prior knowledge* of probability density functions. The use of finite-representation techniques circumvent the major problems in population balance equations and require no input data once a complete single-cell model is formulated.

As currently practiced the finite-representation technique suffers a limitation due to "peak wander". Currently 225 cells represent the population. Within a time increment of 0.01 h only about 10% of the population will undergo division which results in a statistically small sample to be randomized. Due to the small size of the sample, the model predicts slight variations in peak position and shape even under steady-state growth conditions. The stability of the predicted size distribution (or distribution of any other property) can be improved by increasing the number of cells; the improvement would be proportional to the square-root of the total population. An increase to 2,250 computer cells would result in rather stable distributions but also would result in computational requirements that would eliminate most mini computers as vehicles to run such programs.

The problem of "peak wander" is relatively minor in comparison to the engineering benefits. The ability to make *a priori* quantitative predictions of transient behavior of a cell population should be a significant aid to the design engineer. Such systems will allow the development and testing of process control strategies and algorithms without recourse to slow and expensive experimentation. Further, the design of bioreactors in which a population undergoes transient changes in environment (e.g. fed batch) could be expedited by such models.

As Bailey *et al.* (1983) have pointed out the paradigm:

molecular level control systems
↓
single cell kinetics
↓
microbial population dynamics
↓
reactor productivity

is a powerful approach. Models which adopt this approach are ultimately the ones most likely to improve the design and operation of real bioreactors.

SUMMARY

The introduction of chemical structure into models of microbial populations is important in making accurate predictions of a population's biosynthetic capabilities and performance, particularly under transient conditions. The introduction of chemical structure into population models can result in substantial increases in complexity. The modeler must always make the complexity of the model fit the objectives for which the model was constructed. When a detailed knowledge of the distribution of biosynthetic capabilities in a population or when accurate *a priori* prediction of transient responses is required a population model based on a finite-representation technique is the preferred approach. The art of constructing chemically-structured models is young (less than two decades old). Significant opportunities to improve such models exist.

ACKNOWLEDGMENT

The author gratefully acknowledges support by NSF grants 77-20824 and CPE-7921259. In addition the following students have made important contributions to this author's modeling concepts: Chris Dick, Ken Leung, Mike Domach, Mohammad Ataai, and Ann Lee. Program listings are given in appendices in their theses.

NOMENCLATURE

A	extrinsic concentration of limiting nutrient in William's model, M/L^3
A_2	extrinsic concentration of glucose as limiting nutrient in Shuler's model, M/L^3
C_i	intracellular (or intrinsic) concentration of the ith species, M/L^3
D	extrinsic concentration of structural and genetic component in William's model, M/L^3
F	number of replication forks in Shuler's model
k_a	rate constant for ath process, T^{-1}
K_i	saturation parameter for ith species, M/L^3
K_p	saturation-like parameter for feedback inhibition by product, M/L^3
m	total biomass in the system at time t, M
M	extrinsic concentration of total biomass in William's model, M/L^3
M_3	amount of DNA/cell, M
P_3	amount of deoxyribonucleotides per cell, M
r_{ak}	rate of formation of species k by ath process based on biotic volume, M/L^3, T
R	extrinsic concentration of synthetic component in William's Model, M/L^3
t	time, T
V	volume of single cell in Shuler's model, L^3
$\hat{V}$	volume of biomaterial per unit biomass, L^3/M
X_j	mass of jth component per unit mass of biomaterial, M/M

Greek

α	stoichiometric coefficient, M/M
μ	specific growth rate, h^{-1}

REFERENCES

Alberghina, L., Martegani, E., and Mariani, L., Analysis of protein distribution in populations of budding yeast based on a structured model of cell growth. In *Modeling and Control of Biotechnical Processes*, A. Halme, ed., Pergamon Press, p. 33 (1982).

COMPREHENSIVE BIOTECHNOLOGY

The Principles, Applications and Regulations of Biotechnology in Industry, Agriculture and Medicine

EDITOR-IN-CHIEF
MURRAY MOO-YOUNG
University of Waterloo, Ontario, Canada

The Principles of Biotechnology: Engineering Considerations

VOLUME EDITORS
CHARLES L. COONEY
Massachusetts Institute of Technology, Cambridge, MA, USA
and
ARTHUR E. HUMPHREY
Lehigh University, Bethlehem, PA, USA

PERGAMON PRESS
OXFORD · NEW YORK · TORONTO · SYDNEY · FRANKFURT

6

Dynamic Modelling of Fermentation Systems

M. L. SHULER
Cornell University, Ithaca, NY, USA

6.1 INTRODUCTION

A living cell is an immensely complex self-regulated chemical reactor which responds to environmental stimuli (such as changes in nutrient levels, temperature and pH) by altering its internal composition and biosynthetic capabilities. Such changes are not instantaneous but reflect finite time lags in the various biochemical pathways in the cell. Mathematical models that aspire to reflect the basic nature of living organisms must recognize the dynamic nature of such organisms.

Such models are built to fulfill at least one of the following objectives: (1) discrimination among possible mechanisms for the control of cellular processes, (2) bioreactor design and optimization, and (3) process control. The requirements placed on the model building process will differ with respect to the ultimate objective of the model builder.

Typically a model which seeks to be useful in mechanism discrimination at the subcellular level must be very general (and hence complex) and contain a low level of empiricism. Such models must accurately reflect the basic biochemistry of the cell. A high level of detail will invariably require large numbers of parameters; it must be realized that this does not reflect on the validity of a model. A 100-parameter model with no adjustable parameters may be intrinsically more valid than a two-parameter model where both parameters must be adjusted. In models where various subprocesses are self-regulated, and also regulated by the products of other subprocesses, the overall system response may be more dependent on model structure than on the values of the kinetic parameters associated with any individual subsystem.

Such complex models must be closely tied to experimental data to retain validity. Experiments are required for independent parameter estimation and to provide an information base for the formulation of hypotheses about a subcellular control system or pathway. Model predictions incorporating the various hypotheses must be tested against experimental results. Comparison to predictions about the dynamic behavior of the system offers a more stringent test of validity than

does comparison to steady-state experiments. The process is simply that experimental evidence suggests models which lead to testable predictions and to further experiments which lead to refinements in the model resulting in new hypotheses and experiments, *etc*.

The other extreme is the formulation of models solely for process control. In this case the model builder is restricted to variables which can be readily determined on-line. Since the number of variables which can be reliably measured on-line is small (particularly in commercial systems), the model builder will use much simpler models than those intended for mechanism discrimination. Such models generally contain a moderate level of empiricism, particularly when explicit measurements of the product are impossible and productivity must be correlated with other more easily measured parameters. Consequently, models intended for process control will be valid for a relatively narrow operating range of abiotic conditions and will have a minimal number of parameters of which a large fraction may have to be obtained using curve-fitting procedures. The ultimate extreme would be the so-called 'black box' models.

Intermediate between these extremes are models intended to develop a more basic understanding of bioreactor performance or for the actual optimization of a process. Generality is important if a broad range of reactor conditions and types are to be explored and consequently the level of empiricism which can be tolerated is low. Since emphasis is on productivity, the level of biochemical detail required will be tied directly to the nature of the product. Models used for optimization will be mature models already subjected to substantial experimental verification; the results of the optimization undoubtedly require experimental validation but such an experimental program would be less extensive than for either mechanism discrimination or control.

6.2 DEFINITIONS AND IMPLICATIONS

6.2.1 Balanced Growth

Campbell (1957) was probably the first to introduce the term 'balanced growth'. He wrote: '. . . it will be convenient to say that growth is balanced over a time interval if, during that interval, every extensive property of the growing system increases by the same factor'. His definition was based on the behavior of a large population of cells.

Barford and coworkers (1982) have sought to broaden this definition to include the growth of individual cells and sustained oscillations by the culture as a whole. In this chapter we will accept the extension of the definition to an individual cell but not to the case of sustained oscillation in a whole culture since the average concentration of cellular components per unit cell weight would be time dependent. Balanced growth for an individual cell requires that each division cell be an exact replica of the previous cycle.

If a culture is in balanced growth, each individual cell need not be in balanced growth (see the data of Powell, 1958) but on the average a 'typical' cell within the culture will fulfill the definition of balanced growth.

6.2.2 Model Characteristics

Tsuchiya and coworkers (1966) in a pioneering review article suggested a conceptual framework for classifying models of microbial cultures. This framework has been retained, although the terminology has been modified through the years. Harder and Roels (1982) offer a well written summary of distinctions among models.

In this chapter we will concentrate on models which are deterministic rather than probabilistic. A deterministic model allows the exact prediction of future behavior based on specifying the current state vector (essentially values for all variables in the model). Deterministic models become increasingly valid as the number of individual members in the population increase. Generally a total population greater than 10 000 is sufficient to treat the system as deterministic. Special consideration must be given to synchronized or to synchronous cultures where 'all' cells initially divide at the same time and cell number increases in a stepwise fashion. After a few generations asynchrony develops as the distribution of cell division times broadens. Such behavior is deterministic in that with a large cell number the future time course is predictable. However, the development of asynchrony depends upon a random or probabilistic event within a population. Models seeking to simulate such behavior must include some mechanism to recognize such randomness.

Models are generally 'structured' or 'unstructured'. An unstructured model assumes that only a

single variable such as cell number or dry weight is sufficient to describe the biosphere; in essence only the quantity of biomass is important. A structured model allows the division of the biosphere into two or more components. A model which is chemically structured divides the biosphere into chemical components. These components may be real and measurable such as DNA, RNA, protein, *etc.* Alternatively, chemical structure may be imparted with less well-defined components such as 'synthetic component', 'structural component', or similar terms. A model may be non-chemically structured by recognizing that in a pure culture the biosphere consists of cells of different cell sizes and ages and the biosynthetic capabilities of a cell depend on age or size. With a mixed-culture a non-chemically structured model would recognize the existence of different species and would consider the interactions among species. Often the term 'structured model' implies only chemical structure. In this chapter an effort is made to recognize explicitly the two possible forms of structure. In a structured model both quality and quantity of the biosphere are important.

Another distinction arises due to the nature of a microbial culture; it consists of many distinct cells. A 'segregated' or 'corpuscular' model is one that explicitly recognizes that a population consists of individuals each of whom may have distinct properties. A 'non-segregated' or 'distributive' or 'continuum' model does not explicitly recognize the existence of individuals but rather the cell mass is viewed as a lumped biomass which interacts as a whole with its environment.

As long as the properties of interest can be adequately represented by averages, the non-segregated approach is satisfactory. However, if properties with moments higher than first-order are important, then the lack of recognition of the existence of individual cells can be important. For example, suppose that 10% of the total population is responsible for 90% of the product formation. Shifts in the distribution of cell types in the population could be important. With the use of genetically engineered organisms it will be quite possible for a population to contain a wide variety of cell types differing in gene dosages (Imanaka and Aiba, 1981). For such cultures some recognition of segregation in the model will be important.

The mathematical requirements for the non-segregated and segregated models to give identical results have been described (Harder and Roels, 1982; Ramkrishna, 1979). Essentially, the continuum approach can be derived from the segregated approach if: (1) the rate function of a sequence of enzymatic reactions, R, can be factorized out of the probability–density function, and (2) the properties of the cell are statistically independent. Under these conditions it can be demonstrated that the correct formulation of chemically structured, non-segregated models requires the use of intrinsic concentrations (*e.g.* mass of component i per unit mass of total biomaterial) for all biotic components. Abiotic components can be expressed as extrinsic concentrations (*i.e.* component mass per unit of reactor volume). Fredrickson (1976) was the first to articulate this requirement based on physical considerations.

The simplest type of model is unstructured and non-segregated; the Monod equation is an example of such a model. Fredrickson and coworkers (1971) have shown that only structured models can possibly predict the response of a microbial culture in unbalanced growth. Thus the Monod equation can only work under balanced growth conditions. Generally, exponential growth in batch culture and steady-state growth in a single-stage chemostat are considered the only common balanced growth situations. Probably neither exactly fulfills Campbell's definition of balanced growth. Barford and coworkers (1982) cite examples of exponential growth in batch culture which are not balanced. For a truly unstructured model to apply to steady-state chemostat growth, cell composition would have to be the same at all dilution rates; experimental measurements have shown that cell composition varies with dilution rate.

For any transient response structured models must be used. The rest of the chapter will be devoted to models which contain sufficient structure (chemical and/or non-chemical) to be useful in predicting the dynamic response of fermentation systems.

6.3 MODELS OF CELLS IN SUBMERGED CULTURE

6.3.1 Chemically Structured Non-segregated Models

Two of the first chemically structured models proposed were those by Williams (1967) and by Ramkrishna *et al.* (1967). Both were two-component models. Williams (1967) lumped the cell into a synthetic component (primarily RNA) and a structural–genetic component (primarily DNA and protein). Ramkrishna *et al.* (1967) divided the cell into a G-mass (RNA and DNA) and D-mass (proteins). As pointed out by Fredrickson (1976), both models are invalid since intrinsic

Such models have proved useful but are clearly limited since the interactions among members of the population may be essential to maintain the stability of the system.

Another form of structured but non-segregated models is obtained by dividing the total biomass in a mixed culture into components based on species rather than on compositional variables such as RNA, DNA, *etc*. Such an approach makes good sense if each species performs a unique function within the population. It is not necessary that each species component be further divided into compositional categories (although such an approach could be very valuable); the division of the culture into species is sufficient to give the culture 'structure'.

A large number of models for mixed cultures exist. An excellent overview can be obtained from articles by Fredrickson (1977, 1983), Kuenen (1983), and Bazin *et al.* (1983). The classification scheme offered by Fredrickson (1983) is given in Figure 1. Such interactions can give rise to a wide variety of dynamic responses. Even when the manipulated parameters in a chemostat (*e.g.* temperature, feed concentration, flow rate, *etc.*) are held constant, the system may exhibit sustained oscillations. Small perturbations in flow rate or substrate levels in the feed can cause very strong transient responses and possible destabilization of the system. Some examples of the dynamic behavior that can be encountered in steady-flow systems are given in Figure 2.

Mixed cultures are of importance in many natural food fermentations, waste treatment and

Effect of presence of B on growth rate of A	*Effect of presence of A on growth rate of B*	*Qualifying remarks*	*Name of interaction*
−	−	Negative effects caused by removal of resources	COMPETITION
−	0		
−	−	Negative effects caused by production of toxins or inhibitors	ANTAGONISM
−	0		AMENSALISM
−	+	Negative effects caused by production of lytic agents; positive effects caused by solubilization of biomass	EOCRINOLYSIS
+	0	Positive effect caused by production by B (host) of a stimulus for growth of A (commensal) or by removal by B of an inhibitor for growth of A	COMMENSALISM
+	+	See remarks for commensalism. Also presence of both populations not necessary for growth of both	PROTO-COOPERATION
+	+	See remarks for commensalism. Also presence of both populations is necessary for growth of either	MUTUALISM
−	+	B feeds on A	FEEDING (includes predation and suspension-feeding)
−	+	The parasite (B) penetrates the body of its host (A) and therein converts the host's biomaterial or activities into its own	PARASITISM
+	+ (or perhaps 0)	A and B are in physical contact, interaction highly specific	SYMBIOSIS
−	−	Competition for space	CROWDING

Figure 1 Scheme of classification of binary population interactions. The roles of A and B may be reversed. Top part of figure is for indirect interactions, while the bottom is for direct interactions. (From Fredrickson, 1983, with the permission of the American Chemical Society, Washington, DC)

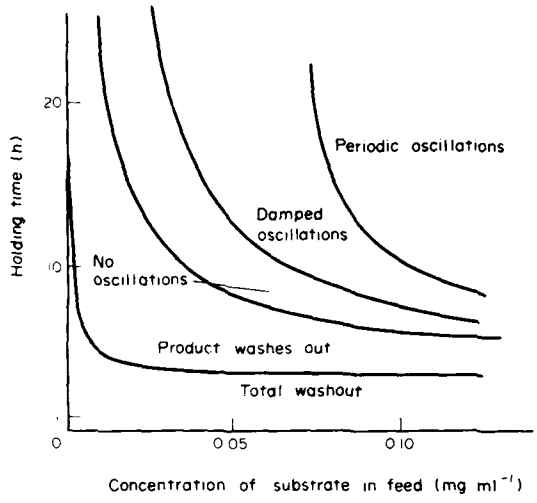


Figure 2 Stability regions of a model for the predator–prey interactions of *Dictyostelium discoideum* and *Escherichia coli* in continuous culture. The model was based on saturation kinetics. Five types of steady-state behavior can be predicted as a function of combinations of holding time and concentration of the limiting nutrient (glucose in this case). The prediction of sustained periodic oscillations was confirmed experimentally with good correlation between the model predictions and experimental data. (From Tsuchiya *et al.*, 1972, with the permission of the American Society for Microbiology)

natural ecosystems. They also represent the case where models of dynamic behavior are essential. Models of populations in which the behavior of each component species is modelled by a chemically structured model have not been accomplished. Such a model would have a much greater potential of truly representing the wide variety of dynamic responses that can be obtained with mixed cultures.

6.3.3 Segregated and Chemically Unstructured Models

Models which are termed segregated but chemically unstructured are based on the presumption that a single variable such as cell age or cell size can completely describe the physiological state of a cell. Thus any cell of say the same size must have the same composition and biosynthetic capabilities. The population model has 'structure' in the sense that the biosynthetic capabilities and composition of the population can be altered as there is a shift in the controlling variable such as size. Such models have the potential to predict transient responses.

Ramkrishna (1979) has summarized a number of aspects of formulating segregated models and reviewed some important aspects of previous studies, however, these models have generally had less impact on biotechnologists working with bioreactors than structured non-segregated models have had.

Shu's (1961) model for product formation is a possible exception. It makes use of an age density function, and product formation is tied to cell age. It is a versatile model and can reproduce the transient profiles typical in a wide variety of fermentations. However, it is difficult to evaluate all the necessary parameters from basic biochemical principles and the model, in practice, has a high degree of empiricism. Such a model may be useful in bioreactor design but not in mechanism discrimination.

As Bailey (1980) has pointed out, the development of segregated models with a significant level of chemical structure has been impeded by the difficulties in obtaining experimental data for model building and verification. The rapid measurement of DNA, protein and RNA (and potentially others) of a single cell can be accomplished with the appropriate fluorescent stains and flow cytometry. The availability of such measurements will undoubtedly act as an impetus to the development of segregated models which allow cells to contain chemical structure.

6.3.4 Population Models Based on Single-cell Models—Segregated and Chemically Structured Models

Models derived from the population point of view, which contain chemical structure as well as recognize segregation, result in equations which are extremely difficult to solve. Shuler *et al.* (1979) described a complex model for the growth of a single cell of *E. coli*. It was suggested that population models containing chemical structure and recognizing segregation might be constructed from a finite-representation technique using each single-cell model to represent some subfraction of the total population.

Nishimura and Bailey (1980) in an important paper starting from the perspective of a single cell of *E. coli* have constructed a model giving analytical solutions for the distributions of cell mass, DNA content, chromosome configuration and total cell numbers. The model requires that the growth rate be specified so that it responds implicitly rather than explicitly to changes in nutrients. The model makes very good predictions of the transient response of such a culture to a shift-up in growth rates. Bailey (1983) has reviewed the use of this general approach to the eukaryotes *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* as well as bacteria.

Shuler and Domach (1983) have reviewed much of the literature concerning the development of models of single cells. Since the number of molecules in a single cell is small, the use of the normal types of kinetic expressions based on concentrations is not strictly allowable. However, if the model cell is to be typical of a large number of cells (at least more than 100), then such kinetic expressions are acceptable. Such an understanding is implicit in almost all of the single-cell models developed.

Shuler and Domach (1983) and Domach *et al.* (1984) have described a complex single-cell model for *E. coli* (see Figure 3). Almost all of the model parameters were estimated from data in the literature. Four parameters associated with cross-wall formation could be evaluated only after the model was run at one growth rate where glucose was rate-limiting. Although the model is complex, it contains only four parameters adjusted within predetermined limits. Such a model provides an ideal framework for the quantitative testing of the plausibility of biological mechanisms. Shuler and Domach (1983) use the model as a basis for testing mechanisms for the control of initiation of chromosome synthesis in *E. coli*.

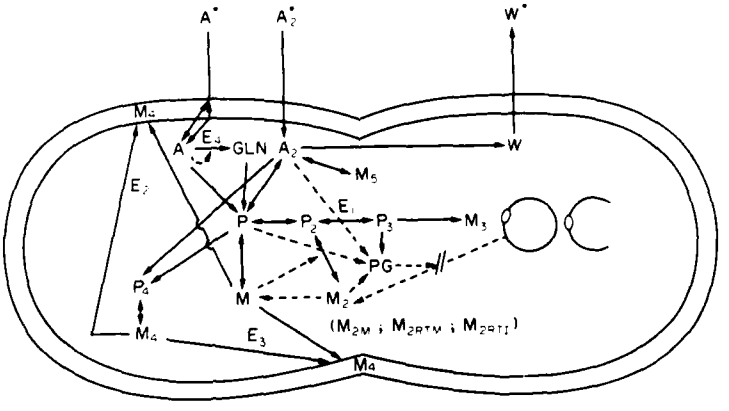


Figure 3 An idealized sketch of the model *E. coli* B/rA growing in a glucose–ammonium salts medium with glucose or ammonia as the limiting nutrient. At the time shown the cell has just completed a round of DNA replication and initiated cross-wall formation and a new round of DNA replication. Solid lines indicate the flow of material, while dashed lines indicate flow of information. The symbols are: A_1, ammonium ion; A_2, glucose (and associated compounds in the cell); W, waste products (CO_2, H_2O and acetate) formed from energy metabolism during aerobic growth; P_1, amino acids; P_2, ribonucleotides; P_3, deoxyribonucleotides; P_4, cell envelope precursors; M_1, protein (both cytoplasmic and envelope); M_{2RTM}, immature 'stable' RNA; M_{2RTI}, mature 'stable' RNA (r-RNA and t-RNA—assume 85% r-RNA throughout); M_{2M}, messenger RNA; M_3, DNA; M_4, non-protein part of cell envelope (assume 16.7% peptidoglycan, 47.6% lipid and 35.7% polysaccharide); M_5, glycogen; PG, ppGpp; E_1, enzymes in the conversion of P_2 to P_3; E_2, E_3, molecules involved in directing cross-wall formation and cell envelope synthesis—the approach used in the prototype model was used here but more recent experimental support is available; GLN, glutamine; E_4, glutamine synthetase; * indicates that the material is present in the external environment. (From Shuler and Domach, 1983, with the permission of the American Chemical Society, Washington, DC)

A population model can be constructed from the single-cell model without the addition of any

adjustable parameters (Domach, 1983; Shuler and Domach, 1983; Domach and Shuler, 1984b). However, a cause for asynchrony must be specified and included in the model; in this case a random variation in the quantity of enzyme responsible for cross-wall formation was chosen (Domach and Shuler, 1984a). Domach and Shuler (1984b) have described the use of such a model for the prediction of the dynamic response of a population of *E. coli* in a single-stage chemostat to a shift in dilution rate. A comparison of experiment to model predictions is given in Figures 4 and 5. Recalling that no adjustable parameters were utilized in developing the population model, the correlation of prediction with experiment is quite remarkable. Thus it appears possible to predict the dynamic response of a large fermenter based solely on basic biochemistry without recourse to empirical expressions. However, such models, while mathematically straightforward, are quite tedious to develop and require substantial computer time.

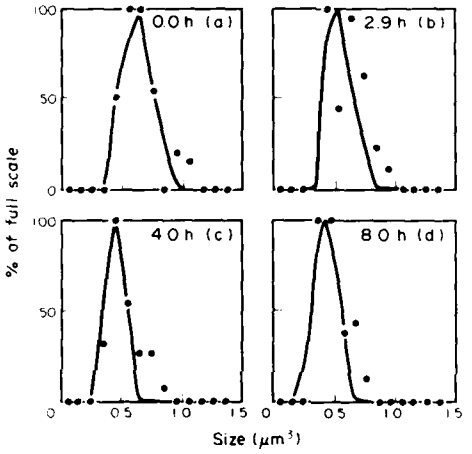


Figure 4 Shift in predicted (●) and observed (—) size distributions due to a flow perturbation in a chemostat. The organism modelled was *Escherichia coli* B rA at 37 °C. Predictions were made using the single-cell model depicted in Figure 3 as a base for a population model using a finite representation scheme. 225 model cells were included in the population scheme; a smoother predicted size distribution would have been obtained if more model cells had been used. Nonetheless, the model accurately predicts the time-dependent shift in cell size and gives a reasonable approximation to the breadth and skew of the size distribution. The initial steady-state distribution is shown in (a) just prior to the decrease in flow. The initial dilution rate of 0.91 h^{-1} was changed to 0.65 h^{-1} at time $t = 0$. (From Domach and Shuler, 1984b)

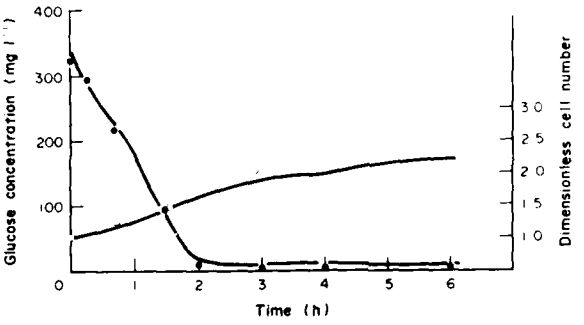


Figure 5 Prediction of transient changes in substrate concentration and dimensionless cell number. The model prediction is denoted by a solid line, while the observed values of substrate concentration and dimensionless cell number are given by ● and ○ respectively. The data are for the experiment described in Figure 4. The model is a population model based on an ensemble of single-cell models as described in Figure 3

6.3.5 Models with Time Delays

Computationally simple models that predict dynamic behavior are particularly desirable for process control. Rather than explicitly introducing a complex kinetic network the effects of cell

adaptation can be included through the use of time delays. Nominally unstructured models modified by inclusion of time delays are potentially promising candidates for making predictions of dynamic behavior.

The underlying rationale for such models can be found in the concept of relaxation times (Harder and Roels, 1982). The concept originated as a means of realistically describing complex thermodynamic systems. A relaxation time characterizes the rate of adaptation of an internal process to changes in the external or abiotic conditions. The system, the biotic phase, is then described in terms of relaxation times and externally observable variables. The smaller the relaxation time the more quickly the internal mechanism adapts to changes in input.

A typical cell is characterized by a large number of processes with widely varying relaxation times, *e.g.* allosteric controls with relaxation times of about 1 s (range 10^{-4} to 10^{2} s) to evolutionary changes with times of 10^{6} s or larger. Not all of these internal processes are usually of importance to the prediction of the behavior of interest. If the rate of change of a variable in the abiotic environment is slow compared to the rate of adaptation of an internal mechanism to that change, then the dynamics of that internal mechanism may be neglected: it will always be at a quasi steady-state with respect to the external variable. In the above example the relaxation time of the internal process is much smaller than a characteristic time associated with the external system. On the other hand, if the relaxation time of the internal process is much larger than the external relaxation time, then that internal process can generally be ignored from a short-term viewpoint such as for process control. For example, the 'normal' dynamic response of a population in continuous culture to perturbations in flow are dissipated in two or three residence times but such changes may have long-term effects in the selection of a subpopulation of cells. Such a selection might not become apparent for many more cell generations.

Consequently, the dynamic behavior of a system could be satisfactorily estimated by only considering those internal processes that have the relaxation times of external changes. These internal processes with smaller relaxation times can be considered to be in a quasi steady-state while those processes with larger relaxation times can be ignored. For process control where the major fluctuations in the abiotic environment can be anticipated, a model recognizing a small number of relaxation times may be quite adequate.

The application of the use of transfer functions to biological systems is an example of the concept of relaxation times. An important example is the model suggested by Young and Bungay (1973). With this model they were able to predict the transient response of a chemostat to perturbations in flow or substrate feed concentrations; parameters predicted were biomass, substrate, protein, RNA and cell number. Results that might be expected are given in Figure 6. The values of the time constants could be estimated from experiments using a 'black box' approach. The application of such techniques in the field of process control is well known (Coughanowr and Kopel, 1965). The essential limitations to this approach are: (1) that only predetermined external variables are changing and at a rate consistent with the experiments to evaluate the time constants, and (2) the transfer approach assumes a linearized system. Since biological systems are highly non-linear, the transfer function will be valid only for relatively small perturbations.

Time delays can also be included in models based on physiological reasons rather than 'black-box' models. Many such models have been recently reviewed by MacDonald (1982). An early example is the discussion of a linear model with discrete delay which was invoked by Finn and Wilson (1953) in considering observations of sustained oscillations of a yeast population in a chemostat. Others have also suggested more complicated expressions making use of not only discrete delays but distributed delays employing a memory function. Such delays may act to approximate the complicated relationship between cell numbers and biomass in a population, or to include the effects of inertial nutrient pools, or to recognize that a cell's previous physiological history will affect its dynamic response to perturbations. Important examples are models as suggested by Powell (1969) which use a memory function to assess the influence of the history of the nutrient concentrations experienced by the population on the population's ability to respond to perturbations. Harder and Roels (1982) have described the use of Powell's model for predicting the specific rate of product formation to specific growth rate.

Models with time delay can usefully simulate a variety of responses. The evaluation of parameters from 'black-box' experiments can provide workable models pertinent to the control of real systems in terms of variables which can be readily measured. While such models are potentially attractive for process control and do have a conceptual justification, they are limited to situations where the potential perturbations are known. The accuracy of the predictions depends on the size of the perturbation of the external variable. Such models are not particularly useful in

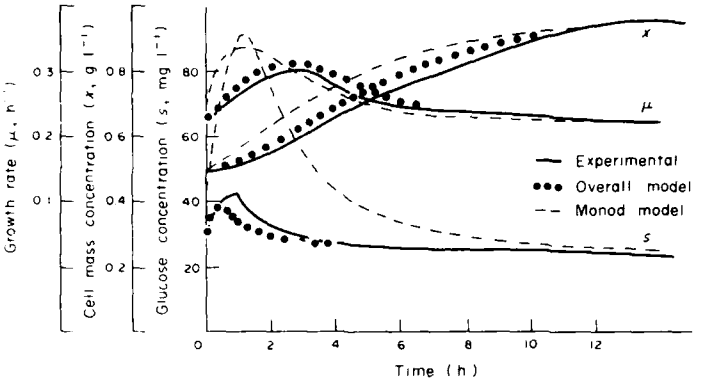


Figure 6 Comparison of predictions from a model derived from a system-analysis perspective, predictions from a Monod model, and experiment. The experimental system was a chemostat for a glucose-limited culture of *Saccharomyces cerevisiae* operating at a dilution rate of 0.20 h^{-1}. In this particular experiment the system was perturbed with a stepwise increase in feed glucose concentration from 1.0 and 2.0 g l^{-1}. x is biomass concentration, μ is growth rate and s is substrate concentration. (From Young and Bungay, 1975, with the permission of *Biotechnol. Bioeng.* and Wiley, New York)

discriminating among hypotheses of how cells function or in predicting the performance of a large variety of bioreactor types employing the given cells.

6.4 MODELS OF CELLS IN SURFACE CULTURES

Although most commercial fermentation processes in the West make use of submerged cultures, surface cultures offer potential advantages (Hesseltine, 1972). In Japan the growth of molds on solid particles (*i.e. the koji process*) is important as a source of enzymes and as a first step in the production of sake. Solid substrate fermentation has been practised successfully on a large scale. Certain mold products, *e.g.* mold spores to be used as insecticides (Miller *et al.*, 1983), require a high level of cellular differentiation and can be best obtained with solid substrate differentiation.

Such a process would, at least at the microscopic level, be always a dynamic one and presents some unique modelling challenges. A colony is always changing; and the system is much more heterogeneous than in submerged culture; spatial considerations cannot be neglected. The difficulties of these challenges coupled with the current low level of commercial activity with solid substrate fermentations has resulted in little real progress in this area. Prosser (1982) has reviewed a number of the suggested models for mold growth on solid substrates. Most models examine a specialized aspect of colony growth such as changes in macromolecular composition during vegetative growth, hyphal tip shape and extension, and growth of individual hyphae. Models for colony formation on solid media exist and are adequate to predict rate of extension of colony and branching patterns. Such models fail to address three important points: cellular differentiation, product formation associated with cellular differentiation, and interaction with nutrients in the solid media.

The rational design of solid substrate fermenters requires dynamic models. Such models would ideally include explicit recognition of the abiotic environment (gas phase and solid medium) and consider differentiation and product formation, and the interaction of colonies through the competition for nutrients or excretion of metabolic byproducts. Mathematically, ordinary differential equations describing mold growth need to be matched to partial differential equations for nutrient (or extracellular byproduct) profiles within the solid media. The macroscopic reactor model would be constructed from the models at the microscopic level.

6.5 SUMMARY

Transient responses of cell populations invariably result in unbalanced growth. Only models that contain structure have the inherent capability of accurately modelling population dynamics.

Structure, in the broadest sense, means that the modeller recognizes that both the quantity and quality of the cell population determines the dynamic behavior of the population. Models vary greatly in complexity and degree of empiricism. The objective that the model is to fulfill determines selection of the model, thus the modeller must be aware of the range of model types and be able to pick a modelling approach matching the desired goal.

ACKNOWLEDGMENTS

The author's work in this area has been supported in part by NSF Grant CPE-7921259.

6.6 REFERENCES

Bailey, J. E. (1980). Biochemical reaction engineering and biochemical reactors. *Chem. Eng. Sci.*, **35**, 1854–1886.

Bailey, J. E. (1983). Single-cell metabolic model determination by analysis of microbial populations. In *Foundations of Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 6, pp. 135–157. American Chemical Society, Washington, DC.

Barford, J. P., N. B. Pamment and R. J. Hall (1982). Lag phases and transients. In *Microbial Population Dynamics*, ed. M. J. Bazin, chap. 3, pp. 55–89. CRC Press, Boca Raton, FL.

Bazin, M. J., C. Curds, A. Dauppe, B. A. Owen and P. T. Saunders (1983). Microbial predation dynamics. In *Foundations of Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 11, pp. 253–264. American Chemical Society, Washington, DC.

Campbell, A. (1957). Synchronization of cell division. *Bacteriol. Rev.*, **21**, 263–272.

Coughanowr, D. R. and L. B. Koppel (1965). *Process Systems Analysis and Control*. McGraw-Hill, New York.

Domach, M. M. (1983). Refinement and Use of a Structured Model of a Single Cell of Escherichia coli for the Description of Ammonia-Limited Growth and Asynchronous Population Dynamics. Ph.D. Thesis, Cornell University, Ithaca, NY.

Domach, M. M. and M. L. Shuler (1984a). Testing of a potential mechanism for *E. coli* temporal cycle imprecision with a structured model. *J. Theor. Biol.*, **106**, 577–585.

Domach, M. M. and M. L. Shuler (1984b). A finite representation model for an asynchronous culture of *E. coli*. *Biotechnol. Bioeng.*, **26**, 877–884.

Domach, M. M., S. K. Leung, R. E. Cahn, G. G. Cocks and M. L. Shuler. (1984). Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/rA. *Biotechnol. Bieng.*, **26**, 203–216.

Finn, R. K. and R. E. Wilson (1953). Population dynamics for a continuous propagator for microorganisms. *J. Agric. Food Chem.*, **2**, 66–69.

Fredrickson, A. G. (1976). Formulation of structured growth models. *Biotechnol. Bioeng.*, **18**, 1481–1486.

Fredrickson, A. G. (1977). Behavior of mixed cultures of microorganisms. *Annu. Rev. Microbiol.*, **31**, 63–87.

Fredrickson, A. G. (1983). Interactions of microbial populations in mixed culture situations. In *Foundations of Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 9, pp. 201–227. American Chemical Society, Washington, DC.

Fredrickson, A. G., D. Ramkrishna and H. M. Tsuchiya (1971). The necessity of including structure in mathematical models of unbalanced microbial growth. *Chem. Eng. Prog. Symp. Ser., No. 108*, **67**, 53–59.

Harder, A. and J. A. Roels (1982). Application of simple structured models in bioengineering. In *Advances in Biochemical Engineering*, ed. A. Friechter, vol. 21, pp. 55–107. Springer-Verlag, New York.

Hesseltine, C. W. (1972). Solid substrate fermentations. *Biotechnol. Bioeng.*, **14**, 517–532.

Imanaka, T. and S. Aiba (1981). A perspective on the application of genetic engineering: stability of recombinant plasmid. *Ann. N. Y. Acad. Sci.*, **369**, 1–14.

Kuenen, J. G. (1983). The role of specialists and generalists in microbial population interactions. In *Foundations of Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 10, pp. 229–251. American Chemical Society, Washington, DC.

MacDonald, N. (1982). Time delays in chemostat models. In *Microbial Population Dynamics*, ed. M. J. Bazin, chap. 2, pp. 33–53. CRC Press, Boca Raton, FL.

Miller, L. K., A. J. Lingg and L. A. Bulla, Jr. (1983). Bacterial, virial, and fungal insecticides. *Science*, **219**, 715–721.

Nishimura, Y. and J. E. Bailey (1980). On the dynamics of Cooper–Helmstetter–Donachie procaryote populations. *Math. Biosci.*, **51**, 305–328.

Powell, E. O. (1958). An outline of the pattern of bacterial generation times. *J. Gen. Microbiol.*, **18**, 382–417.

Powell, E. O. (1969). Transient changes in the growth rate of microorganisms. In *Continuous Cultivation of Microorganisms*, ed. I. Malek, K. Bevan, Z. Fencl, V. Munk, J. Ricica and H. Smrckova, pp. 275. Academic, New York.

Prosser, J. I. (1982). Growth of fungi. In *Microbial Population Dynamics*, ed. M. J. Bazin, chap. 5, pp. 125–166. CRC Press, Boca Raton, FL.

Ramkrishna, D. (1979). Statistical models of cell populations. In *Advances in Biochemical Engineering*, ed. T. K. Ghose, A. Fiechter and N. Blakebrough, vol. 11, pp. 1–47. Springer-Verlag, New York.

Ramkrishna, D. (1983). A cybernetic perspective of microbial growth. In *Foundations of Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 7, pp. 161–178. American Chemical Society, Washington, DC.

Ramkrishna, D., A. G. Fredrickson and H. M. Tsuchiya (1967). Dynamics of microbial propagation: models considering inhibitors and variable cell composition. *Biotechnol. Bioeng.*, **9**, 129–170.

Shu, P. (1961). Mathematical models for the product accumulation in microbial processes. *J. Biochem. Microbiol. Technol. Eng.*, **3**, 95–109.

Shuler, M. L. and M. M. Domach (1983). Mathematical models of the growth of individual cells. Tools for testing biochemi-

mical mechanisms. In *Foundations in Biochemical Engineering, ACS Symposium Series, No. 207*, ed. H. W. Blanch, E. T. Papoutsakis and G. Stephanopoulos, chap. 5, pp. 93–133. American Chemical Society, Washington, DC.

Tsuchiya, H. M., A. G. Fredrickson and R. Aris (1966). Dynamics of microbial cell populations. *Adv. Chem. Eng.*, **6**, 125–206.

Tsuchiya, H. M., J. F. Drake, J. L. Jost and A. G. Fredrickson (1972). Predator–prey interactions of *Dictyostelium discoideum* and *Escherichia coli* in continuous culture. *J. Bacteriol.*, **110**, 1147–1153.

Van Dedem, G. and M. Moo-Young (1973). Cell growth and extracellular enzyme synthesis in fermentations. *Biotechnol. Bioeng.*, **15**, 419–439.

Van Dedem, G. and M. Moo-Young (1975). A model for diauxic growth. *Biotechnol. Bioeng.*, **17**, 1301–1312.

Williams, F. M. (1967). A model of cell growth dynamics. *J. Theor. Biol.*, **15**, 190–207.

Young, T. B., III and H. R. Bungay (1973). Dynamic analysis of a microbial process: a systems engineering approach. *Biotechnol. Bioeng.*, **15**, 377–393.

PLASMID 16, 000–000 (1986)

Mathematical Model for the Control of ColE1 Type Plasmid Replication

M. M. ATAAI[1] AND M. L. SHULER[2]

School of Chemical Engineering, Cornell University, Ithaca, New York 14853

Received January 1986; revised August 1986

A mathematical model for the molecular events controlling replication of ColE1 type plasmids is described. All the model parameters can be evaluated independently. The model simulates plasmid replication and accurately predicts the copy-number of ColE1 plasmids carrying a variety of regulatory mutations. The model is used to test the plausibility of hypotheses concerning the interactions of regulatory elements involved in the replication apparatus. The model favorably supports the mechanism proposed by Tomizawa and co-workers concerning the nature of RNA–RNA interactions and that the Rom protein increases the binding between the two RNA species. The hypothesis that the interactions of RNA I–II increases the susceptibility of RNA II to the action of endonucleases is not a plausible mechanism. © 1986 Academic Press, Inc.

The mechanism by which the copy-number of a plasmid is regulated is of importance to our understanding and use of cells with recombinant DNA. The plasmids containing the ColE1 origin of replication form a particularly important class of plasmids as routine laboratory tools and as key vectors for industrial applications. All of these plasmids need *Escherichia coli* replication enzymes and do not encode any enzymes needed for their own replication (e.g., Tomizawa, 1984). From the practical point of view the key event in normal copy-number control is the control of initiation of replication. These plasmids control copy-number with two inhibitors, "RNA I" and a protein inhibitor.

Our understanding of the process has been greatly increased by recent *in vitro* experiments of Tomizawa and colleagues (Tomizawa, 1984; Tomizawa and Som, 1984). RNA I appears to inhibit the initiation of plasmid DNA replication by preventing the processing of a plasmid transcript (RNA II) to form an RNA primer for the DNA polymerase I. The precursor RNA transcript can hybridize with the template DNA at the origin. The RNA–DNA complex serves as the substrate for RNase H which cleaves the hybridized preprimer RNA to produce an RNA primer. RNA I binds with the complete preprimer transcript and prevents the formation of the stable RNA–DNA complex which is a prerequisite for the formation of the primer RNA. The binding between RNA I and RNA II is a second-order reaction between the two RNA species, and the inhibitor protein [referred to as either Rom (Tomizawa, 1984) or Rop (Cesareni *et al.*, 1982) protein] apparently increases the rate constant for binding between the two RNA species (Tomizawa and Som, 1984). Three hypotheses have been proposed concerning the manner in which the Rom protein could interact with RNA I and RNA II to influence copy-number (Cesareni *et al.*, 1984).

Although the insights gained from these studies are invaluable, they do not allow the a priori prediction of copy-number for a variety of growth conditions and for various mutations in RNA I or II structure. The purpose of this paper is to describe a technique to allow such a priori predictions as well as to quantitatively distinguish among potential hypotheses for subcellular mechanisms such as potential models of interaction of the Rom protein with the two RNA species. Further, the modeling technique described here allows the confirmation of interpretations made from

[1] To whom requests for reprints should be addressed.

[2] Current Address: Department of Chemical Engineering, Polytechnic Institute of New York, 333 Jay St., Brooklyn, N.Y.

0147-619X/86 $3.00

in vitro data by incorporating such parameters in a larger context (i.e., a whole cell) where the complex nonlinear interactions intrinsic to a whole cell are explicitly allowed.

A base mathematical model of *E. coli* B/r suitable for such a study has been described (Shuler and Domach, 1983; Domach *et al.*, 1984; Domach and Shuler, 1984; Ataai and Shuler, 1985a, 1985b; Shuler, 1985). A model of a single cell, representative of a subfraction of a microbial population, can be constructed with significant detail. The model allows explicit interaction of the cell with external nutrients. Population models can be constructed from an ensemble of single-cell models in which each single-cell model represents a subfraction of the total population. The model is able to make accurate predictions of both steady-state and transient changes in cell size, cell composition, growth rates, the timing of initiation of chromosome synthesis, and the length of the C and D periods as a function of external concentrations of glucose and ammonium ions. This explicit interaction with nutrient concentration allows the direct prediction of the effects of various subcellular mechanisms on overall system response—a response that can be easily and quantitatively measured. The use of mathematical models to explore host–plasmid interactions has been pioneered by Lee and Bailey (1984a, 1984b, 1984c) for λdv in *E. coli*. The model of the host cell they used is less robust than the one used here and cannot respond explicitly to nutrient concentrations. Fewer *in vitro* data were available to use to evaluate parameters for control of λdv replication than are currently available for ColE1-type plasmids.

The purpose of this paper is to incorporate models of potential mechanisms for control of plasmid replication into the base model for *E. coli* based solely on independently measured parameters.

The Criteria for Initiation of a Round of Plasmid Replication

The following criteria have to be satisfied for initiation of replication: (1) RNA polymerase binds to the promoter of the RNA II gene and initiates its transcription; (2) The RNA II transcription extends to the origin of replication without binding to an RNA I molecule; (3) The transcript hybridizes with the DNA template at the origin and is processed to serve as a primer for initiation.

To assess the first criterion, the average transcription rate of RNA II promoter ($K_{T_{RNA\,II}}$) is required. If, for example, $K_{T_{RNA\,II}}$ is 10 molecules/h (calculated from Cesareni *et al.*, 1982), it implies that on average a new transcription of RNA II is initiated every 6 min. To assess the second criterion (i.e., whether any of the RNA II transcript would reach the origin of replication before it binds to an RNA I molecule), we must know the time required for the RNA polymerase to reach the origin. The length of this transcript is 555 nucleotides; if the rate of transcription is known, this time can be calculated. Although RNA polymerase occupies about 70 nucleotides (Glass, 1982) it was assumed that transcription of the RNA II has to proceed about 100 nucleotides (also note that RNA I is about 110 nucleotides long) before it can react with an RNA I molecule to avoid steric effects. RNA polymerase travels at a rate of 50–80 nucleotides per second (Molin, 1976; Pace, 1973). Using an average value of transcription rate equal to 65 nucleotides per second, we calculate that the time for RNA polymerase to fully transcribe the RNA II gene is approximately 7 s [i.e., $(555 - 110)/65$].

For the third criterion Tomizawa and Itoh (1982) have reported that approximately half of the transcripts that escape the binding with an RNA I molecule and extend to the origin are capable of serving as a primer for initiation of the plasmid replication.

Description of the Criteria in Mathematical Terms

Description of the first criterion was given in the previous section where $1/K_{T_{RNA\,II}}$ is the average time between two subsequent initiations of transcription from the RNA II promoter and this time can be estimated as 6 min.

For the second criterion we must be able to predict what fraction of RNA II transcripts initiated would escape binding with RNA I in the 7 s from the time of initiation of RNA II transcription.

Binding of RNA II with RNA I is a second-order reaction (Tomizawa, 1984). Thus, the rate of RNA II deactivation due to hybrid formation is

$$\frac{d}{dt}\text{RNA II} = -k_2 \cdot \frac{\text{RNA I}}{\text{VC}} \cdot \text{RNA II} \quad (1)$$

where k_2 is the second-order binding constant and VC is the cytoplasmic volume of the cell. Rearranging Eq. (1) and noting that RNA I concentration is almost constant for the time period of Δt (7 s), then Eq. (1) reduces to

$$\text{RNA II}_{(t=7)} = e^{-k_2 \cdot \text{RNA I/VC} \cdot \Delta t} \cdot \text{RNA II}_{(t=0)} \quad (2)$$

where $\text{RNA II}_{(t=0)}$ is the number of RNA II transcripts initiated, $\text{RNA II}_{(t=7)}$ is the number of unbound RNA II transcripts 7 s after their initiation, and RNA I/VC is the intracellular RNA I concentration. Equation (2) implies that if at any time during the cell cycle, the number of plasmids in which RNA II transciption is initiated is known, and if the values of VC and RNA I at that instant are used, then we can calculate the fraction of RNA II molecules which would extend to the origin without binding with RNA I.

For the third criterion, almost half of the RNA II transcripts which passes the origin of replication without binding with RNA I will not hybridize with the DNA template at the origin or lead to a replication event (Tomizawa and Itoh, 1982). The number of plasmids (PL) after the initiation of transcription of RNA II is obtained from the equation

$$\text{PL}_t = \text{PL}_{t-1} \cdot 0.5e^{-k_2 \cdot \text{RNA I/VC} \cdot \Delta t} + \text{PL}_{t-1} \quad (3)$$

The subscripts $t-1$ and t indicate the values of PL before and after the latest round of plasmid replication, respectively. For example, if the value of the term

$$e^{-k_2 \cdot \text{RNA I/VC} \cdot 7}$$

is equal to 0.2 at a time that transcription from RNA II promoters has been initiated; and if there were 20 plasmids at that time, then $\text{RNA II}_{(t=0)}$ would be equal to 20 molecules. Using Eq. (2) we calculate $\text{RNA II}_{(t=7)} = 20 \times 0.2 = 4$. This illustrates that out of 20 RNA II transcripts initiated, only 4 of them would extend to the origin without binding with RNA I. From Eq. (3) the number of plasmids is obtained. For the example cited above, plasmid copy-number would equal to 22 (i.e., $20 \times 0.5 \times 0.2 + 20$). We assumed a random mode of plasmid replication (i.e., any plasmid at any time has the same probability of serving as a replication template as have any other plasmids).

Other modes of replication are theoretically possible (Rowand, 1969). However, if we assume that a plasmid can replicate only once during the cell cycle, than a model of this form would predict a copy number of two or less which is substantially less than observed experimentally or predicted with random replication (Ataai, 1986). The "Master Copy" hypothesis can be easily dismissed since plasmids of this type can exist in high numbers (Tomizawa and Som, 1984) which would require transciption rates from the RNA II promoter which are far higher than values measured experimentally (Cesareni *et al.*, 1982).

Thus, based on the comparison of model predictions for plasmid copy-number to experiment, the random mode of replication appears to be the only workable hypothesis. Bazaral and Helinski (1970) have previously presented experimental evidence in support of random replication for plasmids with ColE1 origins of replication.

Cell growth continues throughout the cell cycle. After 6 min another round of transcription from the RNA II promoter is initiated. The same procedure is repeated until the cell divides. This step-wise calculation procedure is more computationally convenient than allowing random binding of RNA polymerase to the RNA II promoter. Since the time interval between rounds of initiation of transcription at the RNA II promoter is small compared to the length of the division cycle, this procedure should be a good analogy of the more realistic mode of random RNA II ini-

tiation. At division, equal partitioning of plasmids is assumed.

We could easily incorporate into our cell model any distribution function for the uneven partition of the plasmids between the daughter cells. This has not yet been done to avoid further computational complexity associated with building population models consisting of an ensemble of single-model cells with different plasmid copy numbers and in turn with different growth rates. Also, we have assumed that plasmid DNA replication is nearly instantaneous. This assumption is based on the observations that the DNA content of pBR322 is 4.3 kb. Using an average replication rate for chromosomal DNA and noting that the plasmid replication is unidirectional, it can be calculated that the time it takes to replicate the plasmid DNA is about 10 s.

Model Formulation

To implement the above criteria we need to be able to predict the amounts of RNA I, RNA II, etc. that are available. The binding between RNA I and RNA II can be described as

$$\text{RNA I} + \text{RNA II} \xrightarrow[k_{-2}]{k_2} \text{Hybrid} = \text{inactive complex.} \quad (4)$$

Since k_{-2} is much smaller than k_2, (Tomizawa, 1985) the reaction is essentially irreversible, then

$$\frac{d}{dt}\text{RNA I} = K_{T_{\text{RNA I}}} \cdot \text{PL} - k_2 \frac{\text{RNA I}}{\text{VC}} \times \text{RNA II} - k_{d_{\text{RNA I}}} \cdot \text{RNA I} \quad (5)$$

where $K_{T_{\text{RNA I}}}$ and $k_{d_{\text{RNA I}}}$ are the average transcription and the degradation rate of RNA I. The first term in Eq. (5) is the rate of formation of RNA I. The second term is the rate of deactivation by forming a complex with RNA II, and the last term is the rate of degradation. Similarly,

$$\frac{d}{dt}\text{RNA II} = K_{T_{\text{RNA II}}} \cdot \text{PL} - k_2 \cdot \frac{\text{RNA I}}{\text{VC}} \times \text{RNA II} - k_{d_{\text{RNA II}}} \cdot \text{RNA II}. \quad (6)$$

The amount of the Rom protein made in the cell is calculated from

$$\frac{d}{dt}\text{Rom} = \left(\beta \cdot \text{PL} \cdot 4.0 \times 10^{-18} \cdot \frac{1}{\text{M3}}\right) \cdot \frac{d\text{M1}}{dt} \quad (7)$$

where M1 is the amount of chromosomally encoded protein, M3 is the DNA content of the cell, and 4.0×10^{-18} is the mass of a pBR322 plasmid in grams. In Eq. 7 the rate of Rom protein synthesis is assumed to be proportional to the rate of chromosomally encoded protein synthesis, multiplied by the ratio of the plasmid DNA to the chromosomal DNA. This equation presumes that the level of RNA polymerase is not rate limiting. The proportionality constant for the Rom promoter strength, β, is obtained from the intracellular concentration of the Rom protein in *E. coli* cells carrying inc12rom$^+$, and inc9rom$^-$ mutant plasmids (Tomizawa and Som, 1984).

The enhancement in the binding constant between the two RNA species caused by Rom protein (α), is obtained from

$$\alpha = \frac{\text{Rom/VC}}{\text{Rom/VC} + K_{\text{Rom}}} \cdot \alpha' \quad (8)$$

where α' and K_{Rom} denote the maximum enhancement in the binding constant in the presence of excess Rom protein, and the saturation constant for Rom, respectively. Then, $k_2 = k_2' \cdot (1 + \alpha)$ where k_2 and k_2' represent the binding constant between the RNA species in the presence and absence of excess Rom protein. The values of the model parameters are given in Table 1. The evaluation and justification of these values follows.

Evaluations of the Model Parameters

The degradation rates for RNA I and RNA II are assumed to be the same as that for a typical mRNA. The second-order rate constants for binding of RNA I to RNA II (k_2, k_2') are those reported by Tomizawa (1984) and Tomizawa and Som (1984) which were obtained from *in vitro* experiments. Their

TABLE 1

PARAMETER VALUES FOR MODEL OF ColE1 REPLICATION

Parameter	Value
$k_{d_{RNA\,I}}$	21 h^{-1}
$K_{T_{RNA\,I}}$	63 Transcripts/h promoter
k_2, k_2'	see Table 2
$k_{d_{RNA\,II}}$	21 h^{-1}
$K_{T_{RNA\,II}}$	10 Transcripts/h promoter
k_{Rom}	2×10^{-6} M
l	0.5
β	0.024

measurements were made at 25°C and our simulation is for 37°C. We used a general rule (Q_{10}) for the Arrhenius temperature dependence of binding constants, which effectively results in doubling of the binding rate from 25 to 37°C. K_{Rom} was calculated from the plot of enhancement in binding rate of RNA I–RNA II versus Rom protein concentration (Tomizawa and Som, 1984). The average transcription rate for RNA I and RNA II promoters were estimated using the data reported by Cesareni *et al.* (1984) where they fused the β-galactosidase structural gene to the promoter of RNA I and RNA II genes. These genes were carried by phages and were integrated into the chromosome at a ratio of one-to-one. The activity of β-galactosidase expressed from RNA I and RNA II promoters was 450, and 65 units as defined by Miller (1972). These units of activity can be correlated to the corresponding values of the number of β-galactosidase molecules/cell from which the average transciption rate of these promoters has been estimated. The calculated values of $K_{T_{RNA\,I}}$ and $K_{T_{RNA\,II}}$ were 63 and 10 transcripts/h-promoter, respectively.

Plasmid–Host Interactions

The host is the glucose-limited aerobic single cell model of *E. coli* B/r (Domach *et al.*, 1984). The plasmid contains a ColE1 origin of replication. The host biosynthesis machinery is used to transcribe and translate the genes of the plasmids. The amount of energy consumed for transcription and translation of the plasmid genes is estimated to be the same per bond as for transcription and translation from the chromosome.

The rate equations for the change in amount of precursors (amino acids, ribonucleotides, deoxyribonucleotides) in the host cell model are corrected to include the amount of these precursors used for formation of the plasmid DNA. A separate equation describes the amount of proteins synthesized from the plasmid genes. The rate of formation of plasmid encoded proteins is assumed to be proportional to the rate of chromosomal protein synthesis based on the ratio of plasmid to chromosomal DNA. The rate equation for mRNA formed is modified to include the amount of mRNA formed for synthesis of the plasmid proteins. All of these interactions which occur between the host cell and the plasmids do not affect substantially the host biosysnthesis machinery or growth rate of cells carrying low copy-number plasmids. These interactions become important when considering recombinant plasmids carrying strong promoters or for simulation of high copy-number plasmid mutants. Proteins made in large amounts would cause significant competition for enzymes involved in transcription, translation, and replication and for precursors and energy. The required changes in the computer program are detailed by Ataai (1986).

RESULTS AND DISCUSSIONS

Model predictions of the copy-number of cells carrying different plasmids with a ColE1 origin of replication are compared in Table 2 to the reported experimental values (Tomizawa and Som, 1984). Model predictions compare well with the experiment. Particularly important is the model prediction of decrease in the copy-number as a result of enhancement of the binding constant due to the Rom protein. It should be noted that the simulation is for cells growing in glucose-minimal salt media at the maximum growth rate (i.e., μ = 0.94 h^{-1}), and the experimental results (Tomizawa

Ed Discussion? yes

A8804 T.38,41 t PQSM10 663 5 ME

TABLE 2

COMPARISON OF MODEL PREDICTIONS TO EXPERIMENT[a] FOR VARIOUS COPY-NUMBER MUTANTS

Strain	k_2,K_2 RNA's binding constant $\frac{(ml)}{molecule-h} \times 10^{13}$	Rom	Model predictions		Experiment	
			Copy-number	Decreases	Copy-number	Decreases
pN204	88	−	23	2.3	30	3.0
pNT205	173	+	10		10	
pN212	17	−	125	6.2	200	6.7
pNT214	96	+	20		30	
pNT52	29	−	60	2.9	60	
pNT318	92	+	21		20	3
pNT73	42	−	43	1.6	40	2.0
pNT319[b]	66	+	28		20	
pNT59	35	−	54	2.2	50	2.5
pNT317	78	+	25		20	

[a] Tomizawa and Som (1984).

[b] Mislabeled in original paper as pNT318 (J. Tomizawa, personal communication).

and Som, 1984) are for growth in LB medium at 32°C. Although these two growth situations are quite distinct, the overall growth rate is similar (ca. μ = 1 h^{-1}). Also, both our experimental results and the model predictions indicate only slight variation in copy-number with growth rate for growth rates in the range of 0.3 to 1.0 h^{-1}. The model simulations assume that the transcription rates for RNA II and RNA I are not dependent on growth rate.

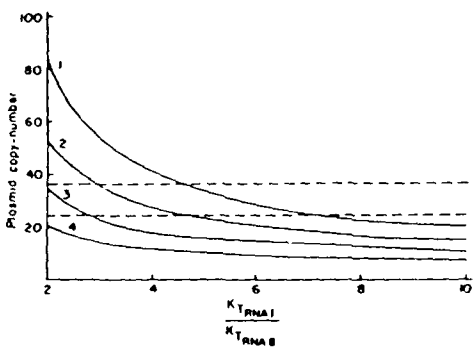


FIG. 1. Effect of variation of RNA I and RNA II promoters strength on plasmid copy-number. The binding constant between the two RNAs is 84×10^{13} (cc/molecule-h). The degradation rate of RNA I and II is 21 h^{-1}. Lines 1, 2, 3, and 4 correspond to $K_{T_{RNA\,I}}$ equal to 45, 90, 180, and 360 transcripts/h-promoter, respectively. The area between the dashed lines correspond to the experimental value of plasmid copy-number where ±20% error in the reported value is assumed.

Changes in transcription rate of RNA I and RNA II with growth rate could lead to stronger variation of copy-number with growth rate. In Fig. 1 we display the predicted copy-number as function of the ratio of transcription rates for RNA I to RNA II. A family of curves is generated as a function of RNA I promoter strength. Runaway replication occurs if $K_{T_{RNA\,I}} \geq K_{T_{RNA\,II}}$. Thus differential changes in $K_{T_{RNA\,I}}$ and $K_{T_{RNA\,II}}$ with growth rate could change copy-number significantly and even if the ratio of $K_{T_{RNA\,I}}/K_{T_{RNA\,II}}$ is maintained at 6.3 a decrease in overall transcription rate with a decrease in growth rate could lead to an increase in copy-number. Other factors have been tested for the sensitivity of the model parameters. Slight variations in assumptions about the frequency of initiation from the RNA–DNA complex do not effect copy-number predictions greatly (22 copies for f = 0.4 and 28 for f = 0.6 while the value of 0.5 used

in Eq. (3) results in a copy-number of 23). If we assume that changes in RNA structure alter susceptibility to degradation by nucleases, then fairly significant changes in copy-number will be predicted (e.g., about 10 at $k_{d_{RNA\,I}}$ = 10.5 h^{-1}, 23 at 21 h^{-1}, and 85 at 42 h^{-1}).

The simulation results in Fig. 1 provide information on the ratio of RNA I and RNA II transcription rates (>1 to prevent runaway replication and > about 2 to minimize fluctuations in plasmid number in cell populations). Even with no prior experimental measurements one can use modeling techniques to place bounds on experimental results. Further, the simulation results provide a basis to predict how plasmid copy-number might be affected if either RNA I or RNA II transcription rates were altered by mutation or if promoters for these two transcripts were replaced by promoters of altered strength.

Since the details of molecular events involved in the control of replication are incorporated into the formulation of the initiation criteria, and all the model parameters are obtained independently, we believe this model can be used to test hypotheses about the interactions of regulatory elements involved in the replication control. As is evident from Table 2, the model is consistent with the finding of Tomizawa and Som (1984) that the Rom protein increases the binding constant between the two RNA species.

Cesareni *et al.* (1984) proposed three hypotheses concerning the ways in which Rom protein could interact with the two RNA species and control copy-number: (1) The interaction changes the secondary structure of RNA II such that it cannot bind to the DNA template at the origin. (2) The interactions cause premature termination of RNA II transcripts and that a potential site for termination exists at a position 220 nucleotides downstream of the promoter of RNA II gene and 20% of transcripts terminate at that point. or (3) Interactions between the RNA's species increase the susceptibility of RNA II to the ribonucleases. The first hypothesis serves as the basis for the model, and the results (Table 2) support the plausibility of this hypothesis. However, for the second hypothesis, the model was used to calculate the fraction of RNA II transcripts of the wild-type plasmid with rom$^-$ mutations which would bind with RNA I at position 220 downstream from the initiation point of RNA II transcripts. It was found that 28% of the precursors initiated would have reacted with RNA I. This result is close to the 20% value observed by Lacatena *et al.* (1984). Thus, either hypothesis 1 or 2 is in quantitative agreement with the experimental observations. However, the model suggests that for the second hypothesis to be plausible a potential termination site must exist at a position 365 or further downstream from the initiation of RNA II transcription. If the termination site at 220 bp found (Lacatena *et al.*, 1984) is the only site, then it is highly unlikely that the second hypothesis is correct. This concept is deduced from the simulation results presented in Table 3; if the last termination site before the origin is at the position 365 bp or upstream of that value, the predicted plasmid copy-number would be too high to be plausible. Similar dependence of copy-number on the position of termination is predicted if the *rom* gene is active.

To investigate the plausibility of the third hypothesis, Eq. (3) was replaced with the equation

$$PL_t = PL_{t-1} + 0.5 \cdot e^{-A_2 \cdot RNA\,I/V_C \cdot \Delta t} \times PL_{t-1} + 0.5 \cdot \text{hybrid} \cdot PL_{t-1} \quad (9)$$

TABLE 3

PLASMID COPY NUMBER VS THE POSITION OF HYPOTHETICAL STRONG TERMINATORS OF RNA I-II COMPLEX

Position of terminator site from the point of initiation of RNA II transcription	Plasmid, Copy-number[a]
545	23
495	31
430	38
365	50
300	65
220	90

[a] The plasmid copy-number measured experimentally is 30.

where the hybrid denotes RNA I–RNA II complex. The last term represents the fraction of hybrid which is not degraded by endonucleases. Equation (9) assumes that either free RNA II or the undegraded RNA II complex can hybridize with the origin and form primer.

The portion of undegraded hybrid is obtained by writing a material balance for the hybrid:

$$\frac{d}{dt}\text{hybrid} = k_2 \cdot \frac{\text{RNA I}}{\text{VC}} \cdot \text{RNA II} - k_{d_{\text{hybrid}}} \cdot \text{hybrid}. \quad (10)$$

Since RNA II = 1 − hybrid, rearrangement of Eq. (4) results in

$$\text{hybrid} = \frac{-k_2 \cdot \text{RNA I}/\text{VC}(e^{-\Delta t(k_2 \cdot \text{RNA I}/\text{VC} + k_{d_{\text{hybrid}}} - 1)})}{k_2 \cdot \text{RNA I}/\text{VC} + k_{d_{\text{hybrid}}}}. \quad (11)$$

In Eq. (11) if $k_{d_{\text{hybrid}}} \to \infty$, then hybrid $\to 0$, which means all the hybrid formed is degraded instantaneously, then Eq. (9) reduces to Eq. (3), and as a result the model prediction would be in accord with the experiment (see Table 2). However, if $k_{d_{\text{hybrid}}} \to 0$, Eq. (9) reduces to

$$\text{PL}_t = \text{PL}_{t-1} + 0.5 \cdot \text{PL}_{t-1} = 1.5 \cdot \text{PL}_{t-1}. \quad (12)$$

Using Eq. (12) instead of Eq. (3) in our simulation, no steady-state plasmid copy-number is achieved, and the copy-number approaches infinity. Thus, if an intermediate value of $k_{d_{\text{hybrid}}}$ is used, a reasonable prediction of copy-number could be made. Those intermediate values were obtained by using Eq. (11) instead of Eq. (3) in the simulation. If the value of k_2 used is for plasmid with wild-type RNAs and either with or without a mutation in the *rom* gene, the corresponding value of $k_{d_{\text{hybrid}}}$ has to be larger than 8000 h^{-1}. This value of the degradation rate for RNA II or the RNA I–II complex is unrealistically high (i.e., half-life of 0.3 s). As a result we believe this hypothesis is not a plausible one.

The ability of this model to test the plausibility of the biochemical mechanisms involved in the regulation of the initiation of plasmid replication, as well as suggesting the conditions under which a hypothesis is plausible, clearly demonstrates the need for constructing mathematical models which are based on the sequence of molecular events which are thought to take place. It should be noted that this type of model is also essential in investigating the problem of genetic stability and finding the optimum bioreactor configuration for maximum production of proteins coded by plasmids.

ACKNOWLEDGMENT

This work was supported in part by ONR Grant N00014-85-K-0580. Portions of this research were conducted on the Production Supercomputer Facility of the Center for Theory and Simulation in Science and Engineering which is funded in part by NSF, New York state and IBM Corporation.

REFERENCES

ATAAI, M. M. (1986). "Mathematical Models for Growth of Single Cell and Population of Escherichia coli under Anaerobic Conditions and Development of a Mechanistic Model for Replication of ColE1 Plasmids." Ph.D. thesis, Cornell University, Ithaca, N.Y.

ATAAI, M. M., AND SHULER, M. L. (1985a). Simulation of the growth pattern of a single-cell of Escherichia coli under anaerobic conditions. *Biotechnol Bioeng* 27, 1051–1055.

ATAAI, M. M., AND SHULER, M. L. (1985b). Simulation of CFSTR through development of a mathematical model for anaerobic growth of *Escherichia coli* cell population. *Biotechnol Bioeng* 27, 1051–1055.

ATAAI, M. M., AND SHULER, M. L. (1986). On the control of ColE1 type plasmid replication. Part II. Hypotheses concerning which plasmids act as templates and sensitivity to alterations in RNA I and RNA II. *Plasmid* 16, 000–000.

BAZARAL, M., AND HELINSKI, D. R. (1970). Replication of a bacterial plasmid and an episome in *Escherichia coli*. *Biochemistry* 9, 399–406.

CESARENI, G., CORNELISSEN, M., LACATENA, R. M., AND CASTAGNOLI, L. (1984). Control of pMB1 replication: Inhibition of primer formation by Rop requires RNA I. *EMBO J.* 3, 1365–1369.

CESARENI, G., MUESSING, M. A., AND POLISKY, P. (1982). Control of ColE1 DNA replication: The rop gene product negatively affects transcription from the replication primer promoter. *Proc. Natl. Acad. Sci. USA* 79, 6313–6317.

DOMACH, M. M., LEUNG, S. K., CAHN, R. E., COCKS, G. G., AND SHULER, M. L. (1984). Computer model for glucose-limited growth of a single cell of Escherichia coli B/r A. *Biotechnol. Bioeng.* 26, 203–216.

DOMACH, M. M., AND SHULER, M. L. (1984). A finite representation model for an asynchronous culture of *E. coli*. *Biotechnol. Bioeng.* 26, 877–884.

GLASS, R. E. (1982). "Gene Function." Univ. of California Press, Berkeley/Los Angeles.

LACATENA, R. M., BANNER, D. W., CASTAGNOLI, L., AND CESARENI, G. (1984). Control of initiation of pMB1 replication: Purified rop protein and RNA I affect primer formation *in vitro*. *Cell* **37**, 1009–1014.

LEE, S. B., AND BAILEY, J. E. (1984a). A mathematical model for λdv plasmid replication: Analysis of a wild-type plasmid. *Plasmid* **11**, 151–165.

LEE, S. B., AND BAILEY, J. E. (1984b). A mathematical model for λdv plasmid replication: Analysis of copy-number mutants. *Plasmid* **11**, 166–~~200~~ 177

LEE, S. B., AND BAILEY, J. E. (1984c). Analysis of growth rate effects on productivity of recombinant Escherichia coli populations using molecular mechanism models. *Biotechnol. Bioeng.* **26**, 66–73.

MILLER, J. H. (1972). "Experiments in Molecular Genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MOLIN, S. (1976). Ribosomal RNA chain elongation rates in Escherichia coli. *In* "Control of Ribosome Synthesis" (O. Maaloe, ed.), Alfred Benzon Symp. 9, pp. 333–339. Academic Press, Orlando, Fla./Copenhagen.

PACE, N. R. (1973). Structure and synthesis of ribosomal ribonucleic acid of prokaryotes. *Bacteriol. Rev.* **37**, 562–603.

ROWAND, R. (1969). Replication of a bacterial episome under relaxed control. *J. Mol. Biol.*, **44**, 387–402.

SHULER, M. L. (1985). On the use of chemically structured models for bioreactors. *Chem. Eng. Commun.* **36**, 161–189.

SHULER, M. L. AND DOMACH, M. M. (1983). Mathematical models of the growth of individual cells. Tools for testing biochemical mechanisms. *In* "Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems" (H. W. Blanch, E. P. Papoutsakis, and G. Stephanapoulos, eds.), pp. 93–133, ACS Symp. Series 207, Amer. Chem. Soc., Washington, D.C.

TOMIZAWA, J. (1984). Control of ColE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* **38**, 861–870.

TOMIZAWA, J. (1985). Control of ColE1 plasmid replication: Initial interaction of RNA I and the primer transcript is reversible. *Cell* **40**, 527–535.

TOMIZAWA, J., AND ITOH, T. (1982). The importance of RNA secondary structure in ColE1 primer formation. *Cell*, **31**, 575–583.

TOMIZAWA, J., AND SOM, T. (1984). Control of ColE1 plasmid replication: Enhancement of binding of RNA I to the primer transcript by the Rom protein. *Cell* **38**, 871–878.

ECL supply ordering page 177

[redacted]

A MATHEMATICAL MODEL FOR PREDICTION OF PLASMID COPY NUMBER AND GENETIC STABILITY IN *Escherichia coli*

by

M.M. Ataai*
and
M.L. Shuler†

School of Chemical Engineering
Cornell University
Ithaca, New York 14853

*Current Address. Department of Chemical Engineering, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, New York 11201

†Corresponding Author

ABSTRACT

The design of bioreactors for genetically modified bacterial cultures would benefit from predictive models. Of particular importance is the interaction of the external environment, cell physiology, and control of plasmid copy-number. We have recently developed a model based on the molecular mechanisms for control of replication of ColE1 type plasmids. The inclusion of the plasmid model into a single-cell *E. coli* model allows the explicit prediction of the interaction of cell physiology and plasmid-encoded functions. The model predictions of the copy number of plasmids with the ColE1 origin of replication carrying a variety of regulatory mutations is very close to that observed experimentally.

All of the model parameters for plasmid replication control can be obtained independently and no adjustable parameters are needed for the plasmid model. In this paper we discuss the model's use in predicting the effect of operating conditions on production of a protein from a plasmid encoded gene and the stability of the recombinant cells in a continuous culture.

INTRODUCTION

The ability to manipulate DNA *in vitro* and then reintroduce the modified DNA into new hosts has greatly expanded the potential role of biological processing. The preferred organism for genetic manipulation is *Escherichia coli* due principally to the wealth of basic knowledge on *E. coli* genetics and physiology. Techniques to introduce new genes into *E. coli* using plasmid vectors and transformation are now routine [1].

Plasmid shedding is a potential hurdle to large-scale processes with recombinant organisms. Several experimental studies have documented plasmid instability [2-6]. Theoretical studies [7,8] have shown that even for batch fermentation on a large-scale those cells losing the plasmid can outgrow the plasmid containing cells leading to a non-productive culture.

Intuitively one might expect that the "metabolic burden" placed on a cell containing a plasmid would reduce a cell's growth rate, particularly in an energy-limited environment when a high-copy-number plasmid is present and plasmid-encoded genes are actively being transcribing. Such intuitive guesses seem to be born out by most experiments [3,5-6]. However, the problem of "metabolic burden" is coupled to problems of plasmid partitioning and stability (e.g. the par locus - 9-13).

One important question is whether this "metabolic burden" is sufficient to predict the growth advantage of revertants over plasmid-containing cells.

The above question is susceptible to analysis using mathematical models. That a need for such models exists and that such models can be formulated is particularly evident in the pioneering papers of Lee & Bailey [14-16] where a model incorporating a mechanism for the control of λdv replication has been formulated.

However, that model is limited by the simplifying assumptions involved in the interaction of the kinetic expressions for plasmid replication and gene expression with host cell functions. The host cell functions do not respond explicitly to changes in the external environment such as concentration of a limiting-nutrient. Thus the model is restricted to the range of growth rates used to establish the empirical expressions and would likely fail under transient growth conditions where substrate concentrations and growth rate become decoupled. Another potential weakness is that this model assumes that the host cell functions are not altered by interactions with the replication of plasmids and expression of r-protein synthesis.

We have developed detailed models of *E. coli* B/r-A [17-21] which could serve as a basis to predict explicitly the effects of plasmid insertion on cell physiology and the interaction of the external environment with the plasmid-containing cell. We have recently extended the single-cell model to include a mechanism of plasmid replication for plasmids using the ColE1 origin of replication [22]. The plasmid pBR322 and its variations use this origin of replication and have proved to be particularly useful vectors in practical systems. Such plasmids use *E. coli* enzymes for replication. Recent articles summarize the details of replication (see 23-33). From the modeling point of view the critical event is the control of initiation of replication. The plasmids control copy-number with two inhibitors: "RNA I" and a small plasmid encoded protein. RNA I acts directly as an inhibitor. A second RNA species, RNA II, is transcribed 550 base pairs upstream of the origin of replication towards the origin and ultimately serves as a primer to initiate plasmid synthesis.

The process is thought to occur as follows: some of the RNA II hybridizes with the template DNA at the origin; the RNA-DNA complex then can serve as the substrate for RNase H, which is thought to cleave the RNA-DNA complex to produce the RNA primer [29]. Under normal conditions the formation of the primer is the rate-controlling step in plasmid replication. RNA I interacts with RNA II during its transcription and prevents the formation of the stable RNA-DNA hybrid which is a prerequisite for the formation of an RNA primer [28,34].

The binding between RNA I and RNA II is a second order reaction. A protein inhibitor has been isolated [35-36]. It now appears that the inhibitor protein (Rom protein) increases the rate constant for binding between the two RNA species [30-31].

Equations which are used to simulate plasmid replication are detailed elsewhere [22] and are summarized in Table 1. The criteria for initiation are summarized in mathematical form in Table 2. The values of the parameters used are given in Table 3 and are justified in reference 22. All of the model parameters were obtained independently (principally from *in vitro* experiments) and were used without adjustment. Definitions of these parameters are provided in the nomenclature section.

The purpose of this paper is to explore how cellular dynamics are altered when this plasmid model is inserted into the framework of a highly structured *E. coli* cell model alters cellular dynamics. In particular we wish to predict the effects of a plasmid induced "metabolic burden" on cellular growth rates, genetic stability in a glucose-limited chemostat, and productivity for production of plasmid-encoded proteins.

MATERIALS AND METHODS

A. Simulation

The host is the glucose-limited aerobic single-cell model of *E. coli* B/r. The plasmid is pBR322. The host biosynthetic machinery is used to transcribe and translate the genes of plasmids. The amount of energy consumed for transcription and translation of the plasmid genes is assumed to be similar to that for transcription and translation from the chromosome.

The rate equations for the precursors (amino acids, ribonucleotides, deoxyribonucleotides) in the host cell-model are corrected to include the amount of these precursors used for formation of the plasmid DNA or plasmid encoded products. The equations for the base model are given elsewhere [17-18]. The Appendix lists only those equations modified from the base model to account for host-plasmid interactions and a separate equation which describes the amount of proteins synthesized by the plasmid genes. The proteins encoded by the plasmid genes, as a first estimate, are assumed to be transcribed at a rate proportional to the rate of protein synthesized from chromosomal DNA. The rate equation for mRNA formed is modified to include the amount of mRNA formed in synthesis of the plasmid proteins. A Runge-Kutta predictor-corrector method is used to solve the set of non-linear ordinary differential equations.

B. Experimental

To test the model predictions a series of continuous culture experiments using a transformant of *E. coli* B/r containing either plasmid pBR322 or p17 were performed. Since plasmids pBR322 and p17 both are derivatives of ColE1 plasmids, use of p17 and pBR322 transformants of *E. coli* B/r allows

direct comparison between the model predictions and experiments concerning the effect of growth rate on plasmid stability and content.

Host: *E. coli* B/r (ATCC 12407)

Plasmids:

The plasmid p17 is a derivative of pBR322 modified by placement of tac promoter [37] upstream of the normal β-lactamase gene. The tac promoter of p17 appears to contain a mutation of the original tac promoter. The tac promoter is a fusion of the lac and trp promoters and can be induced by IPTG [37] which leads to the overproduction of β-lactamase. The plasmid p17 was a gift from Professor D.B. Wilson (Cornell University). Since the frequency of direct transformation into wild type B-strains of *E. coli* is very low, the plasmids were first transformed into *E. coli* WA837 which is restriction$^-$ and methylase$^+$. The plasmid isolated from this strain was used to transform *E. coli* B/r using the standard method described by Maniatis, *et al.* [1]. Since isolated plasmids from WA837 are methylated in the same manner as *E. coli* methylates its chromosome, these plasmids are protected from *E. coli* B restriction endonucleases. The fraction of plasmid-containing cells (stability) was determined by counting the colonies formed in plates with and without ampicilin. The ampicilin concentration of the plates was 40 μg/ml. Each measurement is the average count of 3 to 5 plates using 50-100 μl of sample (3 plates with 100 μl, 1 with 75 and 1 with 50 μl). The plates with too few (⟨20), or too many (⟩ 300), colonies were not counted.

Medium:

The medium was: 0.03 g, $CaCl_2$; 0.01 g, $MnSO_4$; 3 g, K_2HPO_4; 1.5 g, KH_2PO_4; 1.25 g, $(NH_4)_2SO_4$; 0.10 g, $MgSO_4 \cdot 7H_2O$; 10 mg, NaCl; 1.0 mg, $FeSO_4 \cdot 7H_2O$; 37.2 mg, Na_2EDTA, $7H_2O$; and 1.0 g glucose in one liter of distilled water. pH was 6.9 ± 0.1. All chemicals were reagent grade.

β-lactamase assay:

Assays of β-lactamase activity were performed in 50 mM phosphate buffer at pH 7 by using penicilin G (Sigma, St. Louis, MO) as substrate and monitoring the rate of decrease in the absorbance at 240 nm [37]. A unit of activity was defined as 1 μ mole of penicilin G consumed per minute at 25°C.

Continuous culture:

The Bioflo model C-30 (New Brunswick Sci. Co., New Brunswick, NJ) with a working volume of 330 ml was used as a chemostat. The culture medium was introduced through an inlet on top of the vessel. The spent medium with cells was removed through a vented graduated cylinder equipped with a valve, and then to an overflow jar. The flow rate was determined by measuring the time required for the graduated cylinder to be filled with 200 ml of medium. The temperature of the chemostat was maintained at 37°C (± 0.5)

RESULTS AND DISCUSSION

The ability of the model to predict copy-number in *E.coli* cells carrying different ColE1 regulatory mutant plasmids have been compared previously to the experimental values reported by Tomizawa and Som [30] (see reference

[22]). Comparison of model predictions with the experimental values showed an excellent agreement. For example, the model predicts a copy number of 23 with wild-type plasmids when the gene for the production of the Rom protein has been inactivated and 10 in the presence of the Rom protein. The reported experimental values were 30 and 10 respectively. The binding constants were measured at 25°C [31]; our simulation results are for cells growing at 37°C. We assumed an Arrhenius temperature dependence of rate constants which effectively results in a doubling of the binding rate from 25°C to 37°C. Since the details of molecular events involved in control of replication were incorporated into the formulation of the replication mechanism and initiation criteria, and that all the model parameters were obtained independently, the model can be used to test hypotheses about the interactions of regulatory elements involved in the replication control [22].

In Table 4 the dependence of plasmid DNA content on growth rate is shown. As is evident from the simulation results of Table 4, the plasmid DNA content per gram dry weight of cell increases as the growth rate decreases. The increase is due primarily to a decrease in average cell size rather than a change in copy-number. The plasmid copy-number variation over a wide range of growth rate was not significant. The increase in plasmid DNA content with a decrease in growth rate is in agreement with the experimental observation of Siegel and Ryu for the plasmid pPLC-23-trp A1 [39]. However, they also reported a significant variation of plasmid copy-number with growth rate. In another report experimental measurements of pLpl1 plasmid copy-number have shown a plasmid copy-number that is nearly independent of the dilution rate and consequently growth rate [40].

Our measurements of β-lactamase activity (a plasmid coded protein) of *E. coli* B/r/pBR322 host-vector system over a significant range of growth rates are shown in Figure 1. It has been shown that β-lactamase activity and plasmid copy-number are linearly related for plasmids of this type at low to moderate copy number [41]. Thus copy number appears to be independent of the dilution rate for this culture for dilution rates of 0.18 h^{-1} to 0.7 h^{-1}. The model prediction of fairly constant copy-number is based on the assumption that transcription rates of RNA I and RNA II are constant. If transcription rates of RNA I and RNA II vary with growth rate, then copy number will change significantly at low growth rates. However, for either assumption copy-number is nearly constant for dilution rates greater than 0.3 h^{-1}. Copy number would also change significantly if the ratio of RNA I to RNA II varied [22]. At very low growth rates copy number might well increase.

The productivity of plasmid encoded protein synthesis can be predicted using the model. Since copy-number is relatively independent of growth, we find that the assumption that plasmid-encoded protein synthesis is a constant fraction of total protein synthesis results in optimal operation at nearly the same operating conditions as those that maximize biomass productivity (see Table 5). Under such circumstances the optimization problem is simple, and the plasmid protein product can be treated as a primary product as long as the plasmid is stably maintained in the culture. Such dependence of productivity on growth rate has been reported [40,42].

However, a more interesting case is obtained if the plasmid gene is transcribed at a fixed rate ("constitutively") independent of growth rate. Table 6 details the results of such a calculation assuming a transcription rate of 50 transcripts/h-promoter for the plasmid encoded gene. Under such

circumstances the optimal operation conditions are at intermediate dilution rates and product formation can no longer be treated as simply growth associated. The optimum operating conditions depend on promoter strength which can be treated as an independent variable as well as on the parameters (length of cell cycle, cell size, and cell yield) dependent on growth rate. As long as the competition for RNA polymerase or for precursors is not too severe, the assumption of a relatively constant transcription rate from a strong promoter is quite reasonable. However, if we wish to consider effects of even lower growth rates ($< 0.25\ hr^{-1}$) on plasmid encoded protein production, then the possibility of increasing copy number and its effect on protein productivity would need to be considered.

In this productivity calculation we assumed that all of the cells retain their plasmids (i.e. 100% stability). However, cells may lose their plasmids, either through uneven partitioning or physical instability. To find the optimum operational conditions for the maximum productivity of plasmid encoded proteins, the stability of the culture has to be considered. In Table 7 the plasmid stability in a continuous culture of *E. coli* containing the plasmid pBR322rom$^-$ (copy-number ca. 23 is shown). It is assumed that the plasmid gene is fused to different efficient promoters which result in high levels of plasmid-encoded protein. The fraction of the population with the plasmid after m generations was calculated from an equation derived by Seo and Bailey [43]:

$$\Phi_p(m) = \frac{1}{1 + \frac{\theta(1-\eta)}{(1-\theta)(2-\eta)}\left|1 - \left(\frac{2^{\alpha}}{2-\theta}\right)^{m}\right|} \qquad (6)$$

where $\eta=(2-\theta)1/\alpha$, α is the ratio of single-cell growth rate of the plasmid free cells to that of the plasmid-containing cell, and θ is the probability

of the plasmid loss. The value of α was calculated directly from the growth rate of the single-cell model with and without plasmids. Note that the use of single-cell models allows direct calculation of α rather than the indirect method that Seo and Bailey [43] were forced to adopt. θ was calculated from the equation: $\theta = 2^{1-n}$ where n is the plasmid copy-number [40]. The simulation results of Table 7 show that the cells at a higher growth rate (0.94 h^{-1}) are slightly more stable than those at a lower growth rate (0.39 h^{-1}). In these calculations it is assumed that the plasmid instability is only due to uneven partitioning of the plasmids at birth.

Figures 2 and 3 show the experimental results and model prediction of plasmid stability of *E. coli*/B/r/p17 host-vector system in continuous culture at two different dilution rates (0.7 and 0.27 h^{-1}). These cells have the rom gene and copy-number is reduced; consequently the predicted stability is reduced from that predicted in Table 7 for cells which are rom^-. The experimental results are for two identical chemostats which were inoculated with the *E. coli*/B/r/p17 grown in batch in a shake flask. To increase stability during batch growth 70 µg/ml ampicillin was added to both chemostats at the time of inoculation.

Comparing the experimental results in Figures 2 and 3 indicates that the culture is slightly more stable at the higher growth rate ($0.7h^{-1}$), which is in accord with the model prediction. Equation (6) was used in the stability calculation. Use of this equation requires the values of two parameters, α and θ. α was calculated assuming that plasmid coded proteins are 15% of the chromosomal proteins. This value was calculated based on the activity of β-lactamase, which for cells grown at a dilution rate equal to 0.2h h^{-1}, was 37 units/ml. Since 3500 units of activity correspond to 1

mg of β-lactamase, and OD_{600}= 1.4 approximately yields 150 μg protein [44] (OD_{600} of our culture was 1.1), it can be calculated that about 10% of the cell protein is β-lactamase. The β-lactamase gene in p17 is placed downstream of the modified tac promoter, which is controlled in nearly the same manner as the lac promoter. Because of this, the lac IQ gene product is unable to repress the β-lactamase or β-galactosidase gene. This was confirmed by our batch experiments with *E. coli*/B/r/p17 in the presence and absence of IPTG. β-lactamase activity in the culture with IPTG was about 26 units/ml, 10% more than that in the absence of IPTG (26 units vs. 24). We also did batch experiments with *E. coli*/pSKS104. pSKS104 is a pBR322 derivative with insertion of the lac operon in the tetracycline gene of pBR322. The β-galactosidase activity in the culture with and without IPTG was 6900 and 5800 units, respectively. This observation indicates that the presence of plasmid p17 causes expression of lac operon at nearly the fully induced level. β-galactosidase in induced cultures constitutes about 5% of the cell protein [45]. Although β-galactosidase is a chromosomal gene, it is only produced in large amounts in cells with plasmids, which means that in plasmid carrying cells synthesis of about 15% of the cell protein (10% β-lactamase, and 5% β-galactosidase) is due to the presence of plasmids. θ was calculated from the predicted number of plasmids predicted by the model for each growth rate (12 at $D = 0.7\ h^{-1}$, and 11 at $D = 0.3\ h^{-1}$).

Perretti and Bailey [46] have recently extended our base model [17,18] to include more details on control of transcription and translation. This detail will likely become important to predictions of productivity and stability at high levels of expression of plasmid encoded proteins where the availability of RNA polymerase or other key enzymes may become limiting. The model described in this paper which details control of

plasmid number could be integrated with the extended single-cell model developed by Perretti and Bailey [46].

CONCLUSIONS

We have developed a molecular model for replication control of plasmids containing ColE1 origin of replication. The model clearly illustrates the use of a molecular approach in formulating models from the biological mechanisms. Since all the model parameters were obtained independently, the model should prove useful to biologists, to elucidate the nature of interactions involved in the regulation of plasmid copy-number, as well as to engineers, to evaluate the optimum operational conditions for maximum productivity of bioreactors for genetically modified microorganisms. The inclusion of the plasmid model into the highly structured single-cell model allows the prediction of the stability of a recombinant culture under different growth conditions. We believe that this model is the first to be able to make such predictions using only input data on substrate concentration.

ACKNOWLEDGMENTS

This work was supported in part by ONR Grant (N00014-85-K-0580). We are also grateful to Professors Stan Zahler and David Wilson for their assistance in constructing strains.

NOMENCLATURE

$K_{T_{RNA\ I}}$	Transcription rate of RNA I promoter
$K_{T_{RNA\ II}}$	Transcription rate of RNA II promoter t-1
$k_{d_{RNAI}}$	Degradation rate of RNA I
$k_{d_{RNAII}}$	Degradation rate of RNA II
k_2	Second order rate constant between the RNA's species
K_{Rom}	The saturation constant for Rom protein
f	Binding frequency of RNAII transcript with DNA template at the origin
M1	Protein content of a single-cell
M3	Chromosomal DNA
RNAI, RNAII	Number of RNAI and RNAII molecules per cell, respectively
Rom	Number of Rom molecules per cell
VC	Cytoplasmic volume
α	Enhancement in the binding constant caused by Rom protein
α'	Maximum enhancement in the binding constant in the presence of excess Rom protein
β	The proportionality constant

1. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

2. Goldwin, D. and J.M. Slater. 1979. J. Gen. Microbiol., 111:201.

3. Helling, R.B., T. Kinney, and J. Adams. 1981. J. Gen. Microbiol., 123:129.

4. Inselburg, J. 1978. J. Bacteriol., 133:433.

5. Jones, I.M., S.B. Primrose, A. Robinson, and D.C. Elwood. Molec. Gen. Genet., 180:579.

6. Noack, D., M. Roth, G. Muller, K. Undisz, C. Hoftmeier, and S. Casper. 1981. Molec. Gen. Genet., 184:121.

7. Imanka, I. and S. Aiba. 1981. Ann. N.Y. Acad. Sci., 369:1.

8. Ollis, D.F. and H.T. Chang. Biotechnol. Bioeng., 24:2583.

9. Meacock, P.A. and S.N. Cohen. 1980. Cell, 20:529.

10. Nordstrom, K., S. Molin, and H. Aagaard-Hansen. 1980. Plasmid, 4:215.

11. Skogman, G., J. Nilsson, and P. Gustafsson. 1983. Gene, 23:105.

12. Edlin, G., R.C. Tait, and R.L. Rodriguez. 1984. Biotechnol., 2:251.

13. Hartl, D.L., D.E. Dykhuizen, R.D. Miller, L. Green, and J. de Framond. 1983. Cell, 35:503.

14. Lee, S.B. and J.E. Bailey. 1984a. Biotechnol. Bioeng., 26:66.

15. Lee, S.B. and J.E. Bailey. 1984b. *Plasmid*, *11*:151.

16. Lee, S.B. and J.E. Bailey. 1984c. *Plasmid*, *11*, 161.

17. Shuler, M.L. and M.M. Domach. 1983. In Foundations of Biotechnical Engineering, H.W. Blanch, E.T. Papoutsakis, and G. Stephanopoulos, eds. ACS Symp. Ser. 207, Am. Chem. Soc., Washington, D.C., p. 93.

18. Domach, M.M., S.K. Leung, R.E. Cahn, G.G. Cocks, and M.L. Shuler. 1984. Biotechnol. Bioeng., 26:203.

19. Domach, M.M. and M.L. Shuler. 1984b. Biotechnol. Bioeng., 26:877.

20. Ataai, M.M. and M.L. Shuler. 1985a. Biotechnol. Bioeng., 27:1027.

21. Ataai, M.M. and M.L. Shuler. 1985b. Biotechnol. Bioeng., 27:1051.

22. Ataai, M.M. and M.L. Shuler. 1986. *Plasmid* (Submitted)

23. Conrad, S.E. and J.E. Campbell. 1979. Cell, 18:61.

24. Moser, D.R. and J.L. Campbell. 1982. J. Bacteriol., 154:809.

25 Som, T., J. Tomizawa. 1982. Mol. Gen. Genet., 187:375.

26 Selzer, G. and J. Tomizawa. 1982. Proc. Natl. Acad. Sci. (USA), 79:7082.

27 Selzer, G., T. Som, T. Itoh, and J. Tomizawa. 1983. Cell, 32:119.

28. Tomizawa, J. and T. Itoh. 1981. Proc. Natl. Acad. Sci. (USA), 78:6096.

29. Tomizawa, J., T. Itoh, G. Selzer, and T. Som. 1981. Proc. Natl. Acad. Sci. (USA), 78:1421.

30. Tomizawa, J. 1984. Cell, 38:861.

31. Tomizawa, J. and T. Som. 1984. Cell, 38:871.

32. Tomizawa, J. 1985. Cell, 40:527.

33. Cesareni, G., M. Cornelissen, R.M. Lacatena, and L. Castagnoli. 1984. The EMBO J., 3:1365.

34. Muskata, H. and J. Tomizawa. 1984. Cell, 36:513.

35. Cesareni, G., M.A. Muessing, and B. Polisky. 1982. Proc. Natl. Acad. Sci. (USA), 79:6313.

36. Twigg, A.J. and Sherratt, D. 1980. Nature, 283:216.

37. Georgiou, G., J. Chalmers, M.L. Shuler, and D.B. Wilson. 1985. Biotechnology Progress, 1:75.

38. Ataai, M.M. 1986. Ph.D. Thesis. Cornell University, Ithaca, N.Y.

39. Siegel, R. and Dewey D.Y. Ryu. 1985. Biotechnol. Bioeng., 27:28.

40. Koizumi, J., Y. Monden, and S. Aiba. 1985. Biotechnol. Bioeng., 27:721.

41. Bailey, J.E., N.A. DuSilva, S.W. Peretti, J.-H. Seo, and F. Srienc. 1986. *Ann. N.Y. Acad. Sci. 469*:194.

42. Aiba, S. and T. Koizumi. 1984. Biotechnol. Bioeng., 26:1026.

43. Seo, J.H. and J.E. Bailey. 1985. Biotechnol. Bioeng., 27:156.

44. Miller, J.H. 1972. Experiments in Molecular Genetics (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

45. Glass, R.E. 1982. Gene Function (University of California Press, Berkeley, California).

46. Perretti, S. and J.E. Bailey. 1986. Biotechnol. Bioeng. (In Press).

APPENDIX

Equations Changed in Base Model (Ref. 18) to Account for Resource Flow to Plasmid and Plasmid Protein Synthesis

$$\left(\frac{dATP}{dt}\right) = \delta_{A_2}\left(\frac{dA_2}{dt}\right) + \delta_{P_1}\left(\frac{dP_1}{dt}\right)_S + \delta_{P_2}\left(\frac{dP_2}{dt}\right)_S + \delta_{P_4}\left(\frac{dP_4}{dt}\right)_S + \delta_{M_5}\left(\frac{dM_5}{dt}\right)$$

$$+ \delta_{M_1}\left[\left(\frac{dM_1}{dt}\right)_S + \frac{dMPL}{dt}\right] + \delta_{M_2}\left|\left(\frac{dM_{2_{RTI}}}{dt}\right)_S + \left(\frac{dM_{2_M}}{dt}\right)_S\right| + \delta_{M_3}\left(\frac{dM_3}{dt}\right) + \delta_{M_4}\left(\frac{dM_4}{dt}\right)_S$$

$$+ \delta_{PG}\left(\frac{dPG}{dt}\right)_S + \delta_v\left(\frac{dV}{dt}\right) \qquad (3)$$

where $\frac{dMPL}{dt}$ is the rate of protein synthesized from plasmid genes.

$$\left(\frac{dP_1}{dt}\right) = k_1\left(\frac{K_{P_1}}{K_{P_1} + P_1/V}\right)\left(\frac{A_1/V}{K_{P_1A_1} + A_1/V}\right)\left(\frac{A_2/V}{K_{P_1A_2} + A_2/V}\right)$$

$$- k_{TP_1}\frac{K_{TP_1}}{K + A_2/V} \cdot P_1 - \gamma_1\left(\frac{dM_1}{dt}\right) - \varepsilon_2\left(\frac{dP_2}{dt}\right) + \gamma_2\frac{dM_2}{dt} + \varepsilon_3\frac{dP_3}{dt}$$

$$+ \varepsilon_3\gamma_3\left(\frac{dM_3}{dt} + \frac{dPLDNA}{dt}\right) - \varepsilon_4\left(\frac{dP_4}{dt} + \gamma_4\frac{dM_4}{dt}\right) \quad . \qquad (8)$$

where $\frac{dPLDNA}{dt}$ is the rate of plasmid DNA synthesis.

$$\frac{dP_2}{dt} = k_2\left(\frac{K_{P_2}}{K_{P_2} + P_2/V}\right)\left(\frac{P_1/V}{K_{P_2P_1} + P_2/V}\right)\left(\frac{A_2/V}{K_{P_2A_2} + A_2/V}\right)\cdot V$$

$$- k_{TP_2}\left(\frac{K_{P_2}}{K_{TP_2} + A_2/V}\right)P_2 - \gamma_2\left(\left(\frac{dM_{2RTI}}{dt}\right) - \left(\frac{dM_{2RTM}}{dt}\right)_D + \left(\frac{dM_{2M}}{dt}\right)\right)$$

$$- \varepsilon_3\left(\frac{dP_3}{dt} + \gamma_3\left(\frac{dM_3}{dt} + \frac{dPLDNA}{dt}\right)\right) \qquad (9)$$

$$\frac{dP_3}{dt} = k_3\left(\frac{K_{P_3}}{K_{P_3} + P_3/V}\right)\left(\frac{P_2/V}{K_{P_3P_2} + P_2/V}\right)\left(\frac{A_2/V}{K_{P_3A_3} + A_2/V}\right)E_1 - \gamma\left(\frac{dM_3}{dt} + \frac{dPLDNA}{dt}\right) \qquad (10)$$

$$\frac{dM_{2M}}{dt} = \mu_{2_M}\left(\frac{dM_1}{dt} + \frac{dMPL}{dt}\right)_S - k_{TM_{2_M}} \cdot M_{2_M} \qquad (15)$$

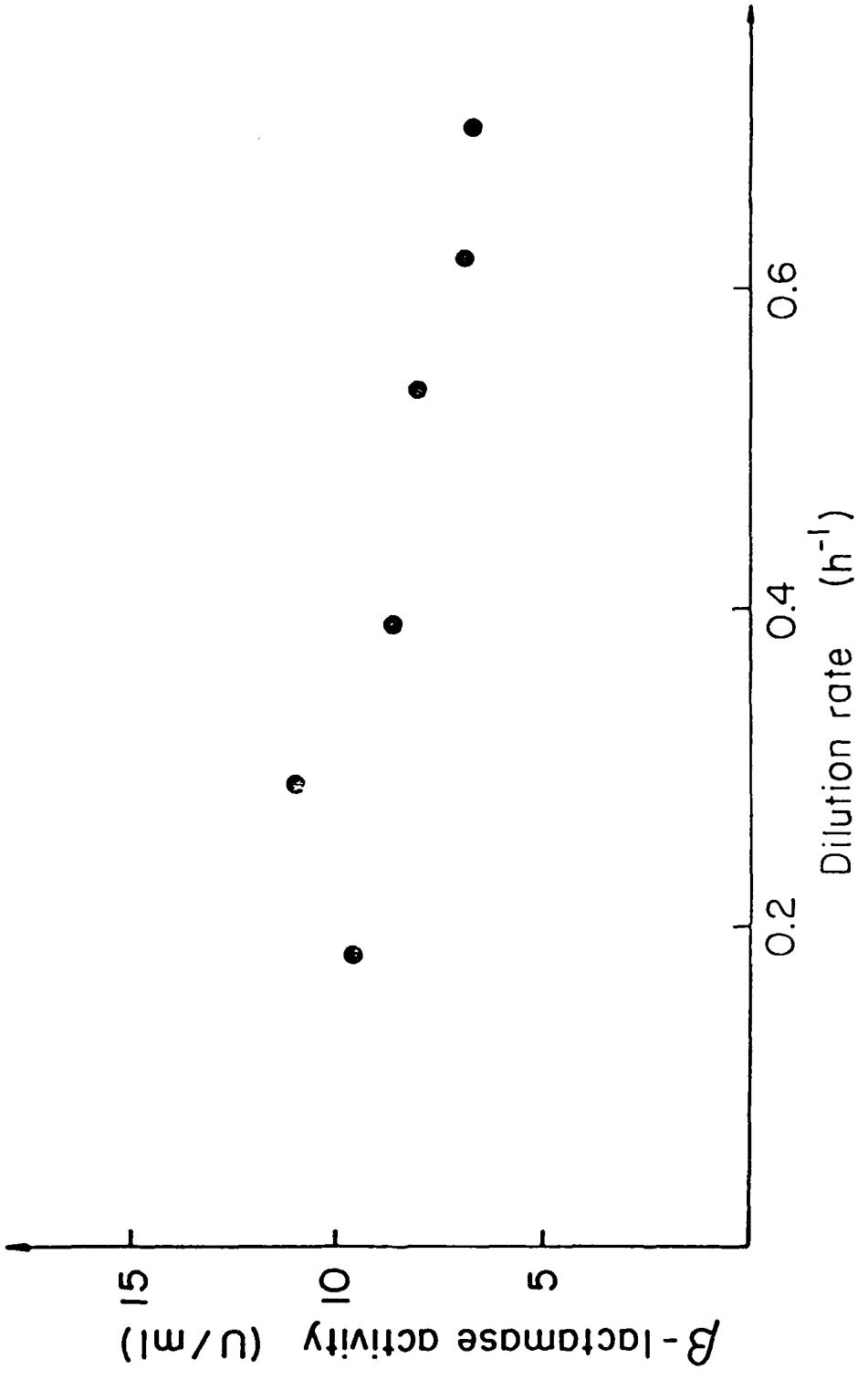
FIGURE LEGENDS

Figure 1. β-lactamase activity from plasmid pBR322 as a function of dilution rate in a continuous culture of E. coli B/r/pBR322.

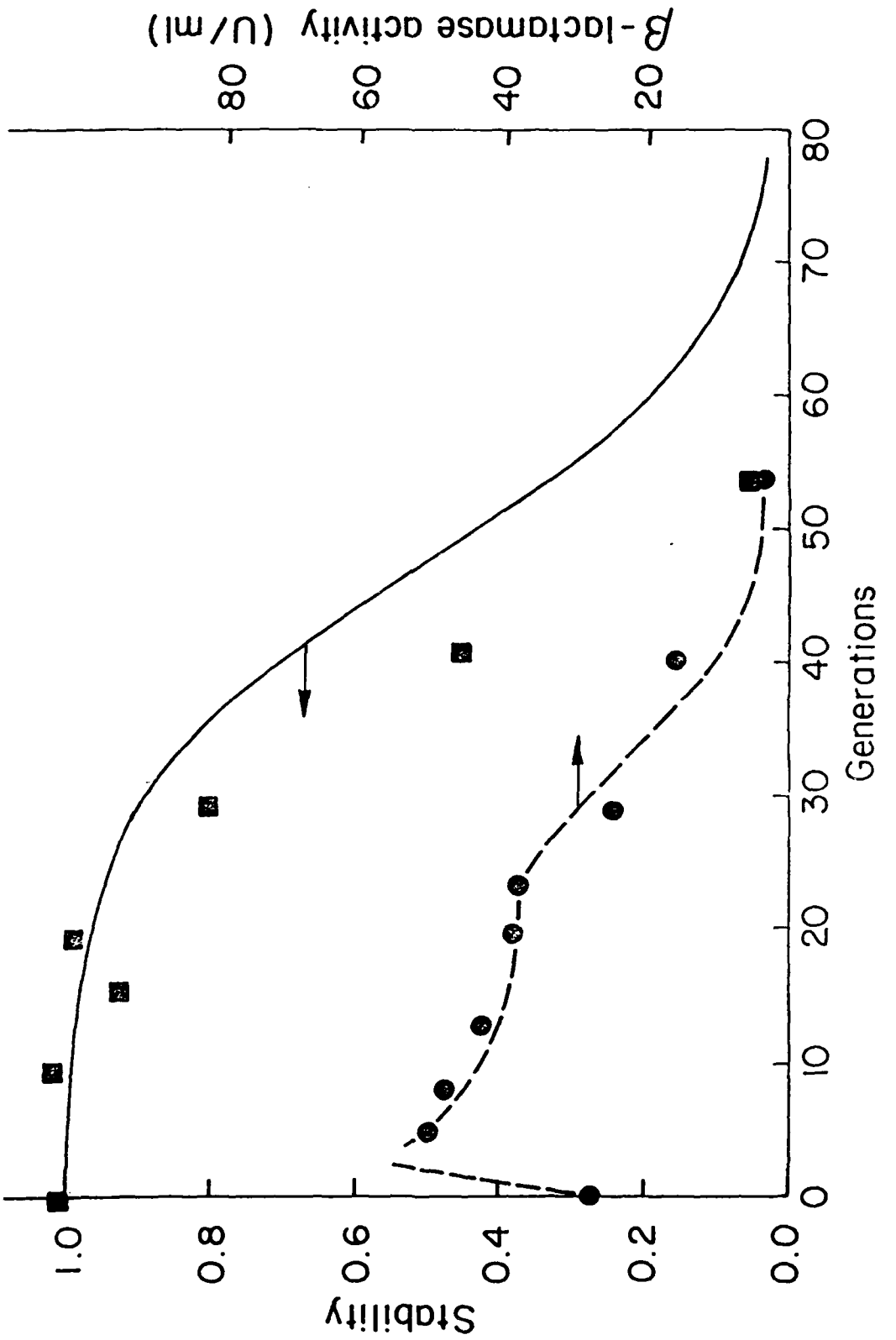
Figure 2. Stability and β-lactamase activity of plasmid p17 in a continuous culture of E. coli B/r/p17. Stability is defined as fraction of plasmid-containing cells in the culture as determined by ability to form macroscopic colonies on ampicilin containing plates. Solid line is the model prediction of stability, where, ■ , denotes our experimental measurement. The, ●, denotes β-lactamase activity where the dashed line is the best line through the experimental points. Dilution rate was 0.7 h^{-1}.

Figure 3. The same as Figure 2 except that the dilution rate is 0.27 h^{-1}.

#1



#2


β-lactamase activity (U/ml)
20
40
60
80
Stability
0.0
0.2
0.4
0.6
0.8
1.0
Generations
0
10
20
30
40
50
60
70
80

#3

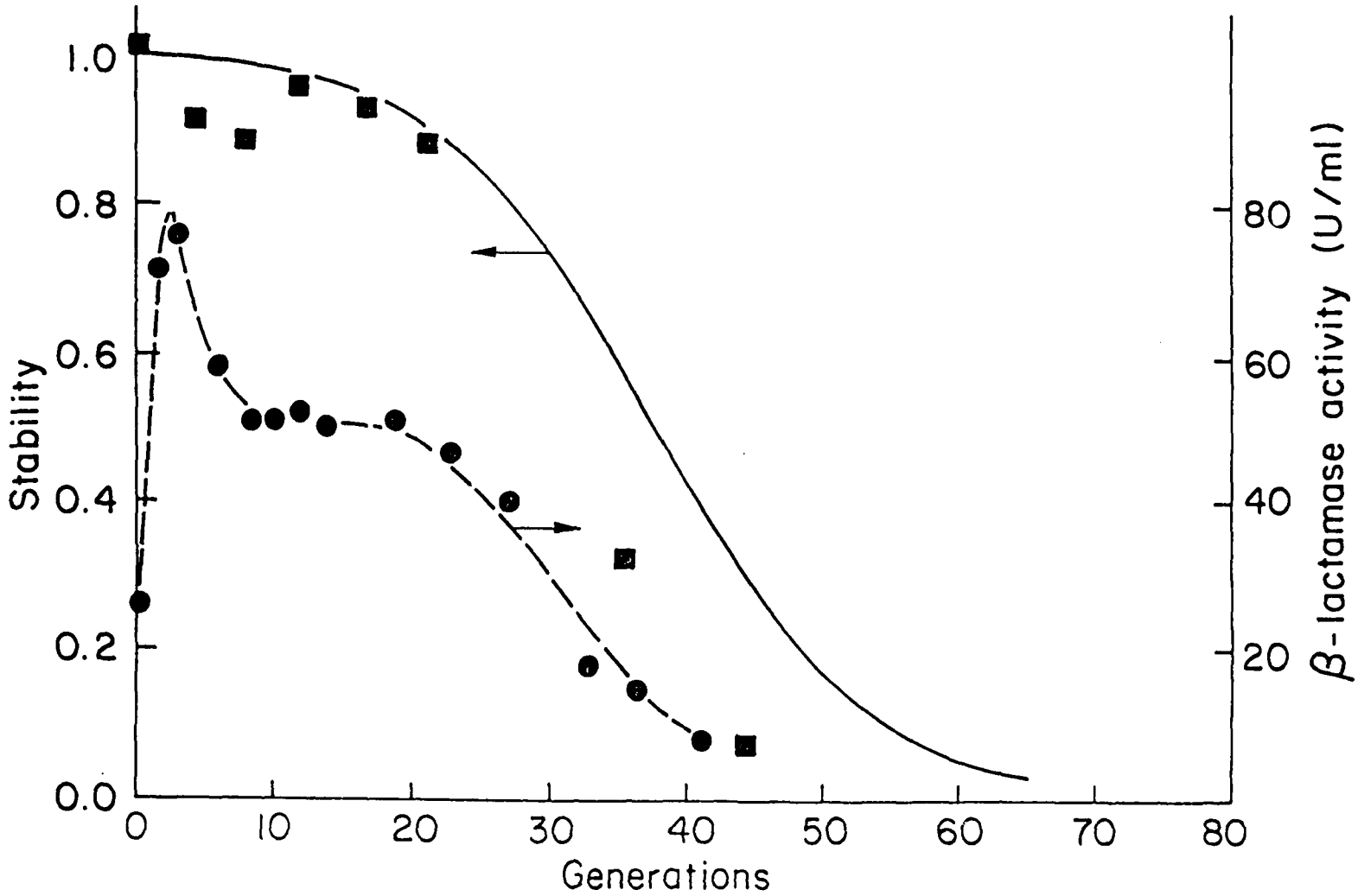

Stability
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
Generations
β-lactamase activity (U/ml)
80
60
40
20

TABLE 1

EQUATIONS WHICH SIMULATE REPLICATION CONTROL OF THE ColE1 PLASMIDS

$$\frac{d}{dt} RNAI = K_{T_{RNAI}} \cdot PL - \frac{k_2}{V_C} \cdot RNAI \cdot RNAII - k_{d_{RNAI}} \cdot RNAI \qquad (1)$$

$$\frac{d}{dt} RNAII = K_{T_{RNAII}} \cdot PL - \frac{k_2}{V_C} \cdot RNAI \cdot RNAII - k_{d_{RNAII}} \cdot RNAII \qquad (2)$$

$$\frac{d}{dt} Rom = (\beta \cdot PL \cdot 4.0 \times 10^{-18} \cdot \frac{1}{M3}) \cdot \frac{dMI}{dt} \qquad (3)$$

where K_T, k_2, and k_d represent the overall transcription rate, second order rate constant, and the degradation rate of the indicated RNA species, respectively. Also,

$$\alpha = \frac{Rom/VC}{\frac{Rom}{VC} + K_{Rom}} \cdot \alpha' \qquad (4)$$

where α is the enhancement in the binding between the RNA's species caused by Rom protein. α' is the maximum enhancement in the binding constant between the RNA's species in the presence of excess Rom protein, and

$$k_2 = k_2'(1+\alpha) \qquad (5)$$

where k_2 and k'_2 represent the binding constant between the RNA's species in the presence and absence of the Rom protein. Thus the model pertains to both plasmids that produce Rom and mutants that do not.

TABLE 2

MATHEMATICAL DESCRIPTION OF THE CRITERIA FOR INITIATION OF A ROUND OF REPLICATION AS DESCRIBED IN THE TEXT

1) The average time between successive binding of an RNA polymerase to RNA II promoter is evaluated from the reciprocal of overall transcription rate of RNA II (i.e. $1/K_{T_{RNA\ II}}$).

2) The number of RNA II molecule which escape binding with an RNA I molecule is calculated by the following equation:

$$\text{RNA II} = \text{RNA II}_o \cdot e^{\frac{-k_2 \cdot \text{RNAI}}{VC} \cdot \Delta t}$$

where Δt is the average time required for RNA polymerase to transcribe RNA II through the origin of replication.

3) The number of plasmids at any time during the cell cycle is calculated from the number of RNA II transcripts initiated (i.e., the first criteria), which escape binding with an RNA I (i.e. second criteria), and hybridize with DNA template at the origin.

$$PL_t = PL_{t-1} \cdot f \cdot e^{-k_2 \cdot \frac{\text{RNAI}}{VC} \cdot \Delta t} + PL_{t-1}$$

where f is the frequency of binding of RNAII transcript to the DNA template at the origin resulting in a DNA-RNAII hybrid which can be cleared by RNaseH. PL_{t-1} and PL_t denote the number of plasmids before and after the latest initiation of transcription from the RNAII promoters.

TABLE 3

THE VALUES OF PARAMETERS*

Parameter	Value
$k_{d_{RNA\ I}}$	21 h^{-1}
$k_{T_{RNA\ I}}$	63 Transcripts/hr promoter
k_2, k'_2	8.8, 17.3 x 10^{-14} $\frac{mL}{molecule-h}$
$k_{d_{RNA\ II}}$	21 h^{-1}
$k_{T_{RNA\ II}}$	10 Transcripts/hr promoter
k_{Rom}	$2 \cdot 10^{-6}$ M
f	0.5
β	0.024

*See reference 22

TABLE 4

EFFECT OF GROWTH RATE ON THE

PLASMID pBR322 rom-DNA CONTENT

Growth Rate	$\frac{\text{mg Plasmid DNA}}{\text{g. bacteria}}$
0.94	0.70
0.67	0.82
0.5	0.92
0.39	0.99
0.29	1.10

TABLE 5

Relative Productivity of Plasmid-Encoded Protein Synthesis as a Function of Growth Rate. Case for Plasmid pBR322 rom$^-$ and the Assumption of Plasmid Protein Synthesis as Constant Fraction of Total Protein Synthesis

Growth Rate h^{-1}	Plasmid Protein Cell $(\frac{g \cdot protein}{cell} \cdot 10^{14})$	Plasmid Protein Productivity $(\frac{g \cdot protein}{ml\text{-}hr} \cdot 10^{5})$*
0.94	0.159	0≠
0.80	0.127	0.270
0.67	0.102	0.215
0.50	0.082	0.145
0.38	0.081	0.116

*Assumes 1.0 g/L glucose in feed.
≠Zero productivity is obtained due to $\mu = \mu_{max}$ and operation at the washout point.

TABLE 6

Productivity of Plasmid-Encoded Protein Synthesis as a Function of Growth Rate

Case for Plasmid pBR 322 rom$^-$ and the Assumption of Constitutive Plasmid Protein Synthesis

Growth Rate h^{-1}	Yield $\frac{g \cdot cells}{g \cdot glucose\ consumed}$	Birth Size μm^3	No. Plasmids at Birth^{++}	No. Plasmid-Encoded Protein Molecules per cell x 10^{5}*	Productivity $\frac{Molecules}{h} \times 10^{-13}$
0.94	0.418	0.453	15	0.405	0≠
0.80	0.410	0.363	14	0.425	8.5
0.67	0.402	0.307	13	0.455	8.9
0.50	0.385	0.255	12.5	0.567	9.6
0.31	0.347	0.211	12.0	0.891	10.0
0.25	0.328	0.200	12.1	1.04	9.6
0.15	0.272	0.177	12.3	1.67	8.8

≠Zero productivity due to cell washout

*Assumes 50 transcripts/h per plasmid, 25 protein molecules per transcript, and plasmid protein degradation at the same rate as host protein.

$^{++}$The number of plasmids at birth could be substantially nigher at dilution rates below 0.30 hr^{-1} if RNAI and RNAII transcription rates vary with growth rate.

TABLE 7

The Effect of Growth Rate and the Level of Expression of the Plasmid Genes on the Stability of a Continuous Culture of Recombinant E. Coli/ pBR322 rom⁻

Glucose Conc. (ppm)	Growth Rate	Plasmid Encoded Proteins / Total Cell Protein	α†	m*
1000	0.940	0.0	1	0
"	0.858	0.15	1.095	308
"	0.762	0.31	1.232	132
"	0.707	0.48	1.328	95
3.27	0.396	0.0	1.0	∞
"	0.350	0.15	1.130	228
"	0.313	0.31	1.265	116
"	0.286	0.48	1.380	78

*number of generations required to reach a population with 10% plasmid containing cells

Growth Behavior and Prediction of Copy Number and Retention of ColE1 Type Plasmids in *E. coli* Under Slow Growth Conditions

by

B.G. Kim
T. A. Good
M. M. Ataai‡
and
M. L. Shuler*

School of Chemical Engineering
Cornell University
Ithaca, NY 14853

*Corresponding Author.

‡Current Address: Department of Chemical Engineering, Polytechnic Institute of New York, Brooklyn, NY 11201

SUMMARY

Changes in plasmid copy number with growth have important implications for productivity of plasmid gene products. Stationary phase cultures often have higher plasmid copy numbers than growing cultures. Further, the use of immobilized cell cultures is of increasing interest and cell growth under these conditions is often very slow. Thus it is of interest to predict the behavior of plasmid-containing cells under slow growth conditions.

We have developed a single-cell model of *E. coli* which contains significant detail on metabolic interactions. Extension to slow growth (doubling times of 20h to 150h) is straightforward. Growth rate predictions for glucose-limited cultures resemble that for the Monrod equation with a maintenance term. Membrane energization is the dominant maintenance energy cost. A minimum glucose concentration is necessary to maintain growth ≥ 0. The value of this minimum depends on cell geometry. Cell "death" (growth < 0) occurs when the cell cannot adjust its size rapidly enough to satisfy the equation:

> Rate Glucose Uptake·Surface Area ≥ Maintenance Energy for Cytoplasmic Functions·Cell Volume + Maintenance Energy for Membrane Energization·Surface Area

Thus cell death depends on initial conditions and the population model can predict which portion of the population will remain viable.

The base single cell model has been extended to include the ColE1 type plasmid replication mechanism. If the transcription rates for the RNAI and RNAII are assumed independent of growth rate, then the predicted plasmid copy number is relatively independent of growth rate even at low growth rates. If the transcription rates for RNAI and RNAII change with growth

rate in a manner directly proportional to the cellular protein synthesis rate, then the copy number is several-fold higher in very slow growing cells although relatively independent of growth rate for moderate growth rates ($\mu > 0.3h^{-1}$).

INTRODUCTION

The normal environment for enteric organisms such as *Escherichia coli* is one of "feast and famine". Over the eons the metabolic control systems for *E. coli* have been tuned to deal with severe environmental fluctuations. This type of selective pressure is part of the reason that metabolic control systems in *E. coli* and in higher eucaryotic cells differ since higher eucaryotic cells are in a relatively more constant environment.

Most fermentation specialists have focused on growth in nutrient rich environments or the "feast" part of the cycle. Yet slow growth or "famine" conditions may be technologically important. The most obvious is an immobilized cell system where cell growth is often intentionally suppressed. Immobilized cell systems offer numerous potential advantages detailed elsewhere in this volume. The potential for cell containment and for control of segregational losses of plasmids make slow growing immobilized cell systems particularly attractive for genetically modified bacteria.

The survival of genetically-modified cells accidentally released from fermentation processes is of regulatory concern. Prediction of survival of a cell population in a "natural" environment is useful in addressing such concerns and certainly requires sound predictive models. In some cases it has been proposed to use genetically-modified cells for the *in situ* treatment of hazardous wastes (1) or in crop protection (2). Such intentional release of cells raises not only regulatory concerns but questions of

efficiency of the treatment. Again models may help to identify those conditions in which the process would be effective without the engineered organisms being displaced too soon by competing natural populations. In all of these circumstances we might anticipate that the cells would be subject to periods of nutrient deprivation.

The consideration of how a cell functions under conditions of nutrient deprivation and slow growth raises intrinsically interesting questions about which biochemical steps are rate influencing.

We have previously described a computer model for a single-cell of *Escherichia coli* (3-8). This model provides a detailed framework in which to explore questions of slow growth and the effects of genetic modification. Unlike many other models this model predicts cellular response (growth rate, composition, size, shape, timing of chromosome replication, etc.) as an explicit function of changes in substrate levels. Currently the model responds to glucose or ammonium ion as limiting nutrients in a glucose-salts minimal medium. Since glucose is both an energy and carbon source, its use as a limiting nutrient provides opportunities for complex behavior.

The purpose of this paper is to describe extension of this model to conditions of slow growth for both wild-type cells and plasmid-containing cells under glucose limiting conditions.

SLOW GROWTH OF WILD-TYPE CELLS

In our previous work we considered growth rates above 0.15 hr^{-1}. In applying the model to low growth rates we discovered an error in the earlier formulation (4). Our intention was to have the rate of amino acid degradation be approximately 50% of the protein degradation rate (for which experimental data existed). Inadvertently the rate constant for amino acid

degradation was set at 50% of the rate constant for protein degradation. However, to have the mass flux for amino acid degradation, the value for the constant, k_{TP1}, needs to be set at 5.0 hr^{-1} since both degradation reactions are assumed to be first-order but the protein pool is much larger than the amino acid pool. This change does not alter model response for growth rates above 0.15 hr^{-1} although it makes a substantial difference at very low growth rates (below 0.05 hr^{-1}). For consistency the degradation rate constant for ribonucleotides, k_{TP2}, was recalculated to be 0.12 hr^{-1} instead of 0.03 hr^{-1}. The change in k_{TP2} has no effect on model response at any growth rate. The saturation constant for energy dependent degradation of amino acids (K_{TP1}) was altered from 11 mg/L to 30 mg/L of A_2 (low molecular organics derived from glucose) to reflect the observed change in A_2 concentration when the cell switched from carbon to energy limitation.

Once these corrections to the original model were made it was possible to use the model to discern which step(s) under glucose limitation alter growth response. Simulation results indicate that at growth rates above 0.15 hr^{-1} the internal A_2 concentration is relatively constant while below 0.15 hr^{-1} the A_2 concentration decreases nearly linearly with growth rate (see Figure 1). We interpret this result to indicate that the cell is primarily limited for carbon above 0.15 hr^{-1} while at lower growth rates the ability to convert A_2 into energy becomes limiting.

One might suppose that if the uptake rate for glucose were increased then the effects of glucose limitation could be suppressed. As shown in Figure 2, simply changing the glucose uptake rate changes the quantitative response of the cell to variations in glucose concentration but not the qualitative response. In each case there is obvious deviation from Monod-type depending on glucose concentration. Further, no matter what uptake

rate is used in these calculations, there is an asymptotic approach to zero growth rate at a finite glucose concentration. The curvature in these plots are suggestive that more than one enzymatic reaction is growth rate influencing (9). Although a similar curvature would be predicted by a model with a diffusional step followed by an enzymatic step (9), diffusion is not allowed in the single-cell model calculations. The concentrations shown in Figure 2 are those that exist at the cell surface and if diffusion is important, these values will be less than the bulk concentration of substrate. Thus, at least two enzymatic-type steps must be controlling in the model at low substrate concentrations.

Clearly the predicted behavior deviates from the Monod equation. We screened several cellular processes to determine if they were responsible for growth rate approaching zero at a relatively high residual glucose concentration. For example, it was found that altering the stringent response (see 10) and the rate of RNA synthesis had no effect on the qualitative nature of the cellular response at very low glucose concentrations although it does at moderate glucose concentrations.

Another possibility is that non-growth associated rather than growth associated phenomena are controlling at low glucose concentrations. As shown in Figure 3, the prediction of a finite residual glucose concentration is simply a consequence of maintenance energy requirements, since removal of these terms results in Monod-like behavior. The model contains maintenance energy terms for precursor and macromolecule turnover, maintenance of membrane energization, and energy spilling reactions such as ppGpp formation and degradation. Membrane energization, which is related to cell surface area, is numerically much larger than the other maintenance terms which are cell volume associated.

An unstructured model where growth rate is described by a Monod-type term plus a maintenance term will give the same qualitative dependence of growth rate on glucose as our more complex single-cell model. Equations of this form are often used for describing the performance of activated-sludge systems, although the interpretation of parameters is more difficult in a multiple species and substrate system.

The single-cell model predictions of growth rate dependency on glucose are consistent with the observations of Daigger and Grady (11) that the RNA limiting theory of growth is inadequate for describing bacterial growth at both intermediate and very low growth rates. However, the prediction of a significant residual glucose concentration was somewhat unexpected since the experiments of Shebata and Marr (12) report the growth of some strains of *E. coli* at glucose concentrations much less than the 0.5 mg/L predicted by the model. However, the presence of a second substrate, such as might arise from cell lysis, can lower apparent threshold concentrations for substrates (13).

Schmidt, *et al.* (14) have developed a method to calculate a threshold concentration of a substrate when simple diffusion to the cell is the limiting step. The calculation of the threshold concentration varies strongly with changes in maintenance energy requirements. We are currently extending the approach of Schmidt, *et al.* (14) to a cell with a hemispherical caps and a cylindrical body. However, based on a cell with a spherical shape we believe that diffusion to the cell could play a significant role (threshold concentrations 0.1 to 1 mg/L) for the high maintenance energy values predicted for *E. coli* B/r at 37°C. These predicted values of maintenance energy have been confirmed experimentally for moderate and higher growth rates (6). However, it is quite possible that membrane

energization costs vary with growth rate since the lipid content of the cytoplasmic membrane changes with growth rate. Harder (in a personal communication in reference 14) believes that maintenance energy may be significantly lower for bacteria growing in nutrient poor environments (10x to 100x). Recently evidence to support decomposing the maintenance term into a constant term and a growth-rate dependent term has been presented (15). By changing membrane composition at lower growth rates the cell may construct a membrane more or less resistant to proton leakage. Since coupling a diffusion model to our cell model is relatively simple, measurements of residual glucose concentrations at low growth rates should provide a means to estimate the maximum allowable value of the maintenance energy and its components.

Thus we compared the model predictions with chemostat experiments. The results are shown in Figure 4 while the techniques used are described at the end of this paper. This comparison is based on the assumption that membrane energization is independent of growth rate and diffusion of glucose to the cell surface is unimportant. Both the data and model predict a finite glucose residual concentration of about 0.5 mg/L or higher. Further, in batch experiments, even well out into the stationary phase, we have measured glucose residuals of about 1.0 to 1.5 mg/L.

The deviations of model prediction from some of the data in Figure 4 is probably due to variations in the viability of the cultures. The prediction derived from the single-cell model assumes 100% viability. Experimental point A which agrees closely with the model has a population of near 100% viability while at point B the population had a viability of 30%. Viability was not measured in the other chemostat experiments. If the viability is less than 100%, dilution rate does not equal growth rate.

(At a dilution rate of 0.05 hr^{-1} with a population that is 30% viable the growth rate of the viable fraction must be 0.17 hr^{-1}).

Agreement of these data with the model cannot totally validate the model. For example, we cannot exclude the possibility of cell lysis which could release a variety of potential substrates. Under conditions of low nutrient availability, *E. coli* no longer uses glucose preferentially (16). Experiments with long residence times can be difficult to conduct due to wall growth or genetic instability although we believe neither of these were present in our case (see "Experimental" section).

In any case the experimental data do suggest multiple steady-states where the viable portion maintains a similar growth rate and substrate concentration but the portion of the population viable changing with dilution rate. Such multiple steady states can probably be predicted by the type of modeling approach employed here.

Consider the following relationship:

$$S \cdot R_{A_2} \geq S \cdot E_M + V \cdot E_T \qquad (1)$$

where S is the surface area of the cell; R_{A_2} is the rate of glucose uptake which is a function of the external concentration of glucose; E_M is the rate of metabolic energy expenditure per unit surface area required to maintain membrane energization (or potential); V is the cell volume; and E_T is the rate of maintenance energy expenditure due to the turnover of cellular constituents. If equation 1 is satisfied, than the cell can remain viable. Equation 1 can be expressed as:

$$(R_{A_2} - E_M) \geq E_T(V/S) \qquad (2)$$

Thus elongated cells with a high S/V ratio are more likely to satisfy equation 2 than short cells. The single-cell model predicts a sharply increasing S/V ratio at low glucose concentrations for cells in a steady-

state situation. Consider, however, the transient situation. A cell growing in a rich medium will have a relatively small S/V ratio. If it is switched to a low glucose concentration, can it change its shape sufficiently rapidly to satisfy equation 2? If we had a population of cells distributed over a wide variety of cell sizes and S/V ratios, we could anticipate that some cells could adjust to a large shift of glucose concentrations while others would not. Thus the apparent steady-state obtained after four or five residence times would be a function of the initial size distribution of the population. Work is currently underway to confirm this possibility.

The model is numerically stable even at very low glucose concentrations. At a glucose surface concentration of 0.55 mg/L the model predicts a doubling time of 140 hr while at 0.50 mg/L cell death is predicted since the internal glucose concentration is negative.

These results are with wild-type cells. When cells are actively expressing plasmid encoded genes for non-essential proteins one would expect even greater sensitivity to low glucose concentrations since the plasmid-encoded functions can be like an extra maintenance energy term.

EFFECTS OF SLOW GROWTH ON PLASMID-CONTAINING CELLS

We recently have reported the extension of the single-cell model to incorporate a mechanistic model for the replication of plasmids with the ColE1 origin of replication (17,18). Such plasmids are commonly used as laboratory or industrial vectors with pBR322 and its derivatives being examples.

The mechanism for ColE1 replication has been well explored by Tomizawa and colleagues (19,20). The plasmid encodes information for two species of

RNA. RNAII, whose transcription is initiated 555 base pairs upstream from the origin of replication, can combine with the origin of replication on the plasmid. The RNA-DNA complex serves as the substrate for RNaseH which cleaves the hybridized preprimer RNA to produce the RNA primer necessary to initiate replication. RNAI is a shorter transcript (about 110 base pairs long) transcribed in the opposite direction from RNAII, but on an overlapping section of the plasmid. RNAI acts as an inhibitor of replication. RNAI binds to RNAII and prevents the formation of the stable RNA-DNA complex which is a prerequisite for the formation of primer RNA. The binding between RNAI and RNAII is a second-order reaction; a plasmid encoded protein, Rom, can increase the binding rate.

The model incorporates this mechanism and can make reasonable estimates of plasmid copy-number for a variety of mutations to genes for RNAI or RNAII which alter their binding rates and consequently the degree of inhibition (17). The equations are described elsewhere in detail (17). Briefly the model requires all plasmids to initiate rounds of RNAII transcription at the same time. The transcription rate for RNAII determines the time interval between application of the equation used to calcualte the number of new plasmids formed a pre-existing set.

In this equation the intracellular concentration of RNAI will strongly alter plasmid copy number by altering the probability that an RNAII transcription event will result in initiation of replication. The concentration of the Rom protein alters the value of the binding rate constant and thus influences copy number but not as strongly as RNAI.

Using the model we have found that if the RNAI transcription rate is less than that for RNAII that runaway replication will occur. Clearly specific inducible promoters coupled to the regions encoding for RNAI or

RNAII could be used to manipulate plasmid copy number and changes in copy number rather precisely.

The mechanistic model for ColE1 replication was tested for predictions of copy number for various mutations in RNAI and RNAII in the presence or absence of the Rom protein and for conditions of moderate nutrient limitation. In extending the model to conditions of lower growth rates to determine the effect of growth rate on copy number it is necessary to consider the effects of growth rate on the transcription of RNAI and RNAII. Two possibilities have been considered by us. First is that the transcription rate of RNAI and RNAII are not changed by growth rate. The second is that both transcription rates vary as the overall transcription rate of the cell varies.

The implications of these two assumptions are shown in Figures 5 and 6. The difference in predicted copy number between these two assumptions is insignificant for growth rates above about 0.30 hr^{-1}. However, the assumption of growth dependent transcription rates leads to a rapid increase in copy number at very low growth rates. If the transcription rate is assumed independent of growth rate, the copy number would change very little with growth rate.

The growth dependent transcription rate certainly seems intuitively reasonable. The prediction of increased plasmid copy number under nutrient starvation is consistent with a number of published reports. Steuber and Bujard (21) report a several fold increase in copy number as cells enter the stationary phase. However, interpretation of their results are complicated by the use of spectinomycin to alter growth rate.

Siegel and Ryu (22) have reported on the effects of nutrient limitation on copy number and plasmid content of a plasmid, pPLc23-trpA1, in *E.*

coli M72. This plasmid has the ColE1 origin of replication. In the above system the plasmid content (mgDNA/g.bacteria) increased 8.0 fold when the growth rate was decreased from μ = 1.12hr^{-1} to 0.06 hr.$^{-1}$ The model predicts a 8.2 fold increase in plasmid content going from μ = 0.09 hr^{-1} to 0.93 hr^{-1}. The plasmid content predicted by the model is higher than measured by Siegel and Ryu (ca. 0.3 mg/g versus 0.5 mg/g at = 0.9 hr^{-1})but this difference reflects differences in plasmid, host cell size, and maximum growth rates. Siegel and Ryu measured 32 plasmids per cell at μ = 1.12 hr^{-1} and 74 at μ = 0.06 hr^{-1}. The model predicts copy number changing from 20 at μ = 0.93 hr^{-1} to 64 at μ = 0.09 hr^{-1}. *E. coli* B/r's maximum growth rate in minimmal medium is 0.93 hr-1 which is slower than *E. coli* M72's maximum growth rate. Thus the model predictions assuming variable transcription rates for RNAI and RNAII give the same relative increase in copy number with decreasing growth rate.

Clearly if the promoter for RNAI was effected differently from that for RNAII as a function of intracellular composition (i.e. growth rate) then copy number dependence on growth rate would be altered from that predicted here.

Once copy number is known it is possible to predict the depression of growth rate due to expression of plasmid-encoded protein synthesis. If copy number and the growth rate depression due to plasmid-encoded protein synthesis is known, then a simple formula derived by Seo and Bailey (23) can be applied to predict the stability of the culture. We have done so with this model for moderate growth rates (μ = 0.3 hr^{-1} and 0.7 hr^{-1}) and have been able to predict the number of generations before 10% of the population contains plasmids within 35% when compared to our experimental

data (18). At the lower growth rate fewer generations are required before instability is observed.

We believe that this type of analysis will aid in designing reactors with cyclic or other dynamic operating patterns. In particular, an increase of copy number at low growth rates suggests the possibility of improving plasmid-encoded protein production by cycling the culture through a "feast-famine" cycle. Further, by extending the model to lower growth rates we expect to have a basis for the design and rational understanding of immobilized cell systems using genetically-engineered *E. coli*.

ACKNOWLEDGMENT

We gratefully acknowledge support, in part, by ONR GRANT (N00014-85-K-0580). This research was conducted on the Production Supercomputer Facility of the Center for Theory and Simulation in Science and Engineering, which is funded, in part, by the National Science Foundation, New York State and IBM Corporation.

REFERENCES

1. Powledge, T. M. 1983. Biotechnol. 1:743.
2. Anonymous. 1985. Biotechnol. 3:109.
3. Shuler, M. L., S. K. Leung & C. C. Dick. 1979. Ann. N Y. Acad. Sci. 326:35.
4 Domach, M. M, S. K. Leung, R. E. Cahn, G. C. Cocks, & M. L. Shuler. 1984. Biotechnol. Bioeng., 26:203.
5 Shuler, M. L. & M. M. Domach. 1983. In Foundations of Biotechnical Engineering, H. W. Blanch, E. T. Papoutsakis, and G. Stephenopoulos, eds. ACS Symp. Ser. 207, Am. Chem. Soc., Washington, D.C., p.93.
6. Domach, M. M. & M. L. Shuler. 1984. Biotechnol. Bioeng. 26:877.
7 Attai, M. M. & M. L. Shuler. 1985. Biotechnol. Bioeng. 27:1027.
8 Shuler, M. L. 1985. Chemical Eng. Commun. 36:161.
9. Dabes, J. N., R. K. Finn & C. R. Wilke. 1973. Biotechnol. Bioeng. 15:1159.
10. Gallant, J. 1979. Ann. Rev. Genetics 13:393.
11. Daigger, G. T. & C. P. L. Grady. 1982. Biotechnol. Bioeng. 26:1427.
12. Shebata, T. E. and A. C. Marr. 1971. J. Bacteriol. 107:210.
13. Law, A. T. & D. K. Button, 1977. J. Bacteriol. 129:115.
14. Schmidt, S. K., M. Alexander & M. L. Shuler. 1985. J. Theor. Biol. 114:1.
15. Pirt, S. J. 1982. Arch. Microbiol. 133:300.
16. Harder, W. & L. Dijkhuizen. 1982. Philosophical Trans. Royal Soc. London, Ser. B. 297:459.
17. M. M. Ataai & M. L. Shuler. 1986. Plasmid (in press).
18. M. M. Ataai & M. L. Shuler. 1987. Biotechnol. Bioeng. (accepted).
19. Tomizawa, J. 1984. Cell 38:861.

20. Tomizawa, J. & T. Som. 1984. Cell 38:871.

21. Stueber, D. & H. Bujard. 1982. EMBO J. 1:1399.

22. Siegel, R. & D. D. Y. Ryu. 1985. Biotechnol. Bioeng. 27:28.

23. Seo, J. H. & J. E. Bailey. 1985. Biotechnol. Bioeng. 27:156.

APPENDIX

Experimental Methods

Organism

E. coli B/r-A (ATCC 12407) was obtained from the American Type Culture Collection, Rockville, Maryland and used in all experiments.

Medium

A modified "C" medium was used for chemostat experiments, batch experiments, and culture maintenance. The medium contained 3.00g K_2HPO_4, 1.50g KH_2PO_4, 1.25g $(NH_4)_2SO_4$, 0.10g $MgSO_4 \cdot 7H_2O$, 0.01g NaCl, 0.001g $FeSO_4 \cdot 7H_2O$, 0.0372g $Na_2EDTA \cdot 7H_2O$, 0.1g $CaCl_2$, 0.01g $MnSO_4$, 1.0g glucose and one liter of distilled water. The pH was 7.0 ± 0.1. All chemicals were reagent grade. The medium was sterilized in three parts, $CaCl_2$ and $MnSO_4$, glucose, and the remaining salts to prevent carmelization of the glucose and precipitation of calcium salts.

Glucose Assay

Glucose concentration was measured enzymatically using the glucose S.V.R. kit (Calbiochem, LaJolla, Ca.). The kit directions state to dilute the enzyme with 15.0 ml of distilled water and then use 3.0 ml of enzyme solution and 0.02 ml of sample. To measure very low concentrations of glucose (down to 0.5 ± 0.2 mg/L) the directions were modified as follows. The enzyme was diluted with 5.0 ml of distilled water. 1.0 ml of the enzyme solution was then added to 3.0 ml of sample. The absorbance was measured at 340 nm. For low glucose concentrations the calibration is linear and of the form:

$$[glucose] = \frac{\Delta A}{0.0345}\ mg/L$$

where ΔA is the difference in absorbance between the prepared sample and a blank containing 1.0 ml of enzyme solution and 3.0 ml of distilled water. The presence of media salts did not effect the calibration.

The medium sample was prepared by withdrawing about 5.0 ml from the growth chamber and rapidly filtering through a 0.45 νm Millipore filter attached to a syringe.

Cell Concentration

Cell concentration was monitored by measuring the absorbance of the cell suspension at 600 nm.

Cell Viability

The percent viable cells was determined by measuring the optical density of a sample and then plating 10^{-5} to 10^{-7} dilutions and counting colonies. This was compared to colonies/ml measured at the same optical density for cells growing in exponential phase.

Continuous Culture

The chemostats used were Bioflow Model C30 (New Brunswick, Edison, N.J.) with working volumes ranging from 320 ml to 360 ml. A peristallic pump (Econo-Column pump by BioRad, Richmond, Ca.) feed system using silicon tubing was employed. The feed rate could be modified to give dilution rates ranging from 0.025 hr^{-1} to 0.15 hr^{-1}. The system was mixed with a small impeller and sparged with sterile, humidified air. Air was supplied using a diaphram pump (Optima by Hagen, Mansfield, Mass.) and passing the air through a packed column filled with water maintained at 37°C. Air was filtered using a Millipore cartridge filter. The chemostat was maintained at 37±0.5°C. No pH control was used. The steady state pH in all experiments was 6.7 ± 0.05.

Nominal residence times were determined by measuring the working volume of the reactor at the appropriate air supply and stirring rates and dividing by the flow rate determined by measuring the time required to pump out a 10 ml pipet. Several residence time distribution studies were performed using red food coloring as a tracer for residence times ranging from 8 hr to 40 hr under conditions similar to actual chemostat operation. This showed that the nominal residence time was within five percent of the actual residence time measured.

The medium was steam sterilized in 10 liter glass jars for 40 minutes as described earlier. The chemostat was inoculated with about 10 ml of a 12 hr shake flask culture of *E. coli*. Batch growth was followed to insure a maximum growth rate of at least 0.9 hr^{-1}. After an optical density of 1.0 was reached the reactor was set for continuous flow.

Steady state was determined by monitoring the glucose concentration and the cell density within the chemostat. Steady state was assumed if at least six residence times had lapsed and cell density and glucose concentration remained constant or two consecutive residence times.

FIGURE LEGENDS

1. The model predicts a nearly linear decrease in internal concentration of low molecular organics ("glucose") at low growth rates (below 0.15 hr^{-1}) while this internal concentration is relatively independent of growth rate for moderate growth rates (0.15 hr^{-1} to 0.9 hr^{-1}).

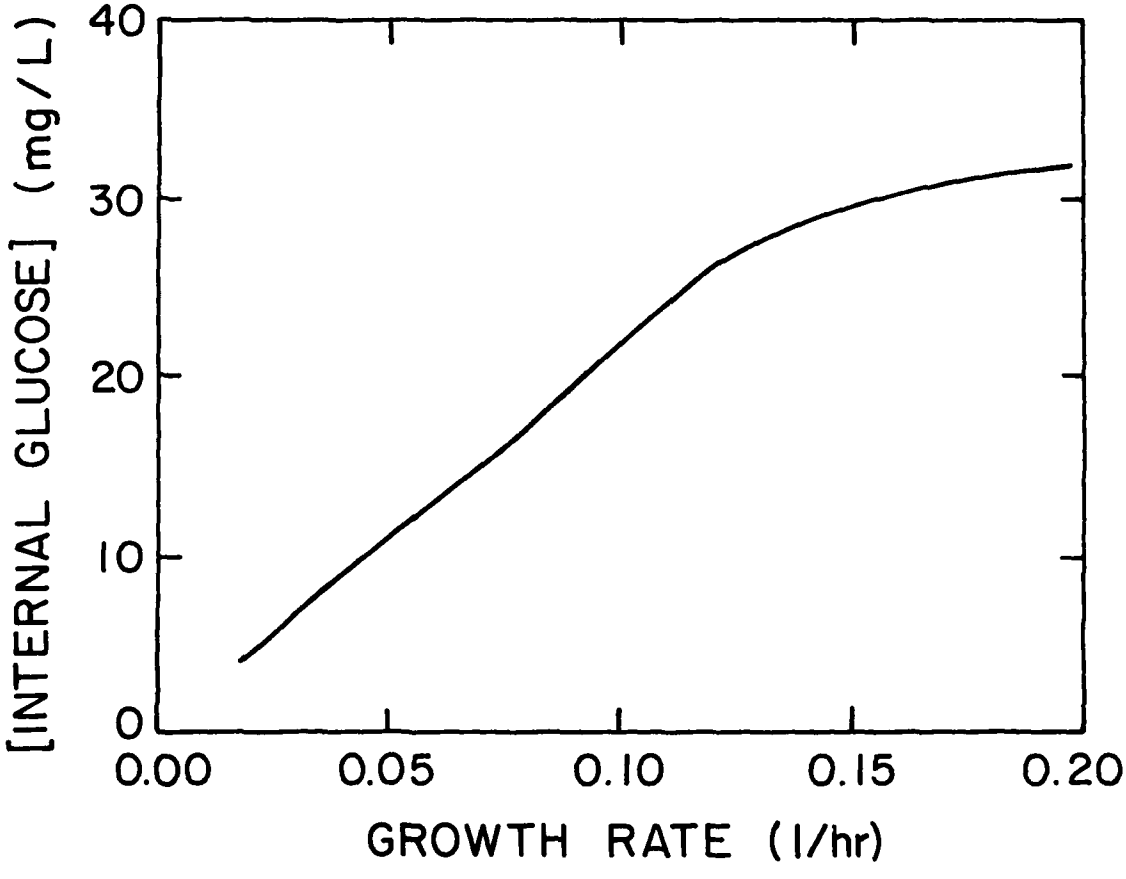
2. Altering the maximum rate of glucose uptake does not alter the qualitative nature of the predicted growth response to changes in substrate concentration. The dash/dotted, dashed, solid and dotted lines represent the response if the maximum rate of glucose uptake is set at 50%, 75%, 100% and 125% of its normal value (ref. 4), respectively.

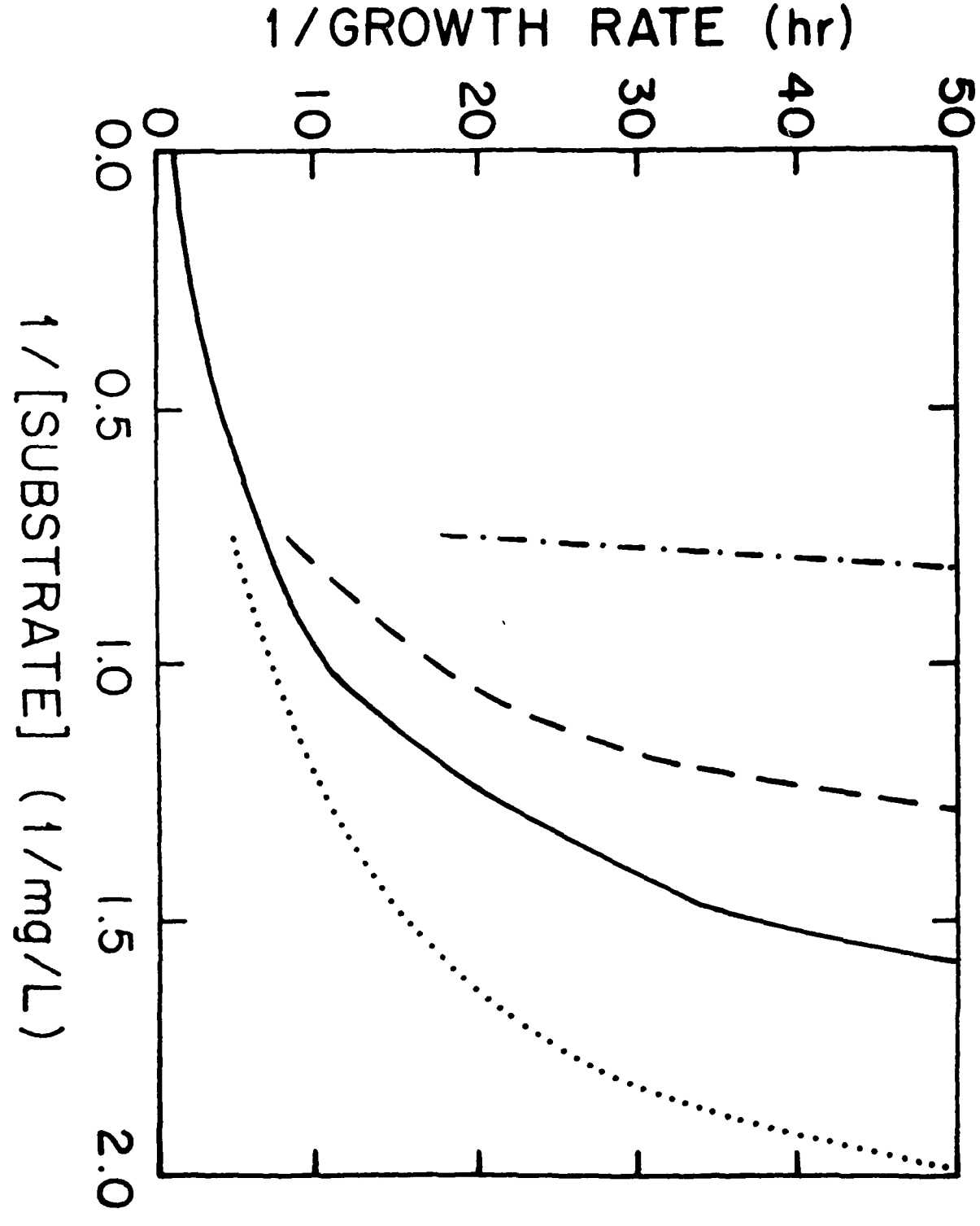
3. By removing the maintenance energy terms the model predicts Monod-like dependence of growth rate on substrate. The dashed line represents the prediction from the complete model. The solid line is the model prediction when maintenance terms are removed. The dotted line represents the Monod equation with K_S equal to 4 mg/L.

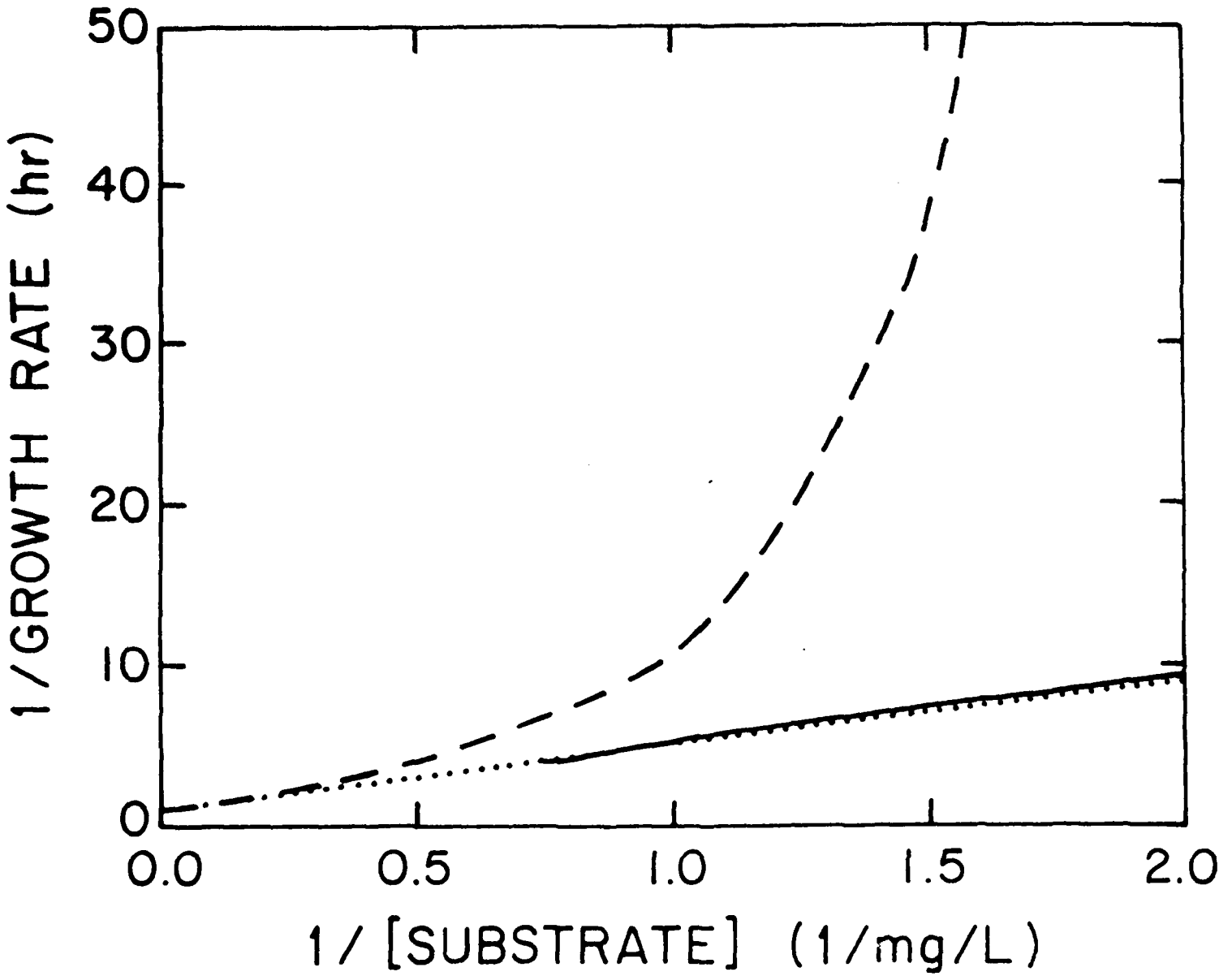
4. The predicted and measured dependence of residual glucose concentration on dilution rate is shown. Dilution rate equals growth rate if and only if all cells are viable. The solid line represents predictions of the model. For measurement A culture viability was 100% while at measurement B only 30% of the cells were viable.

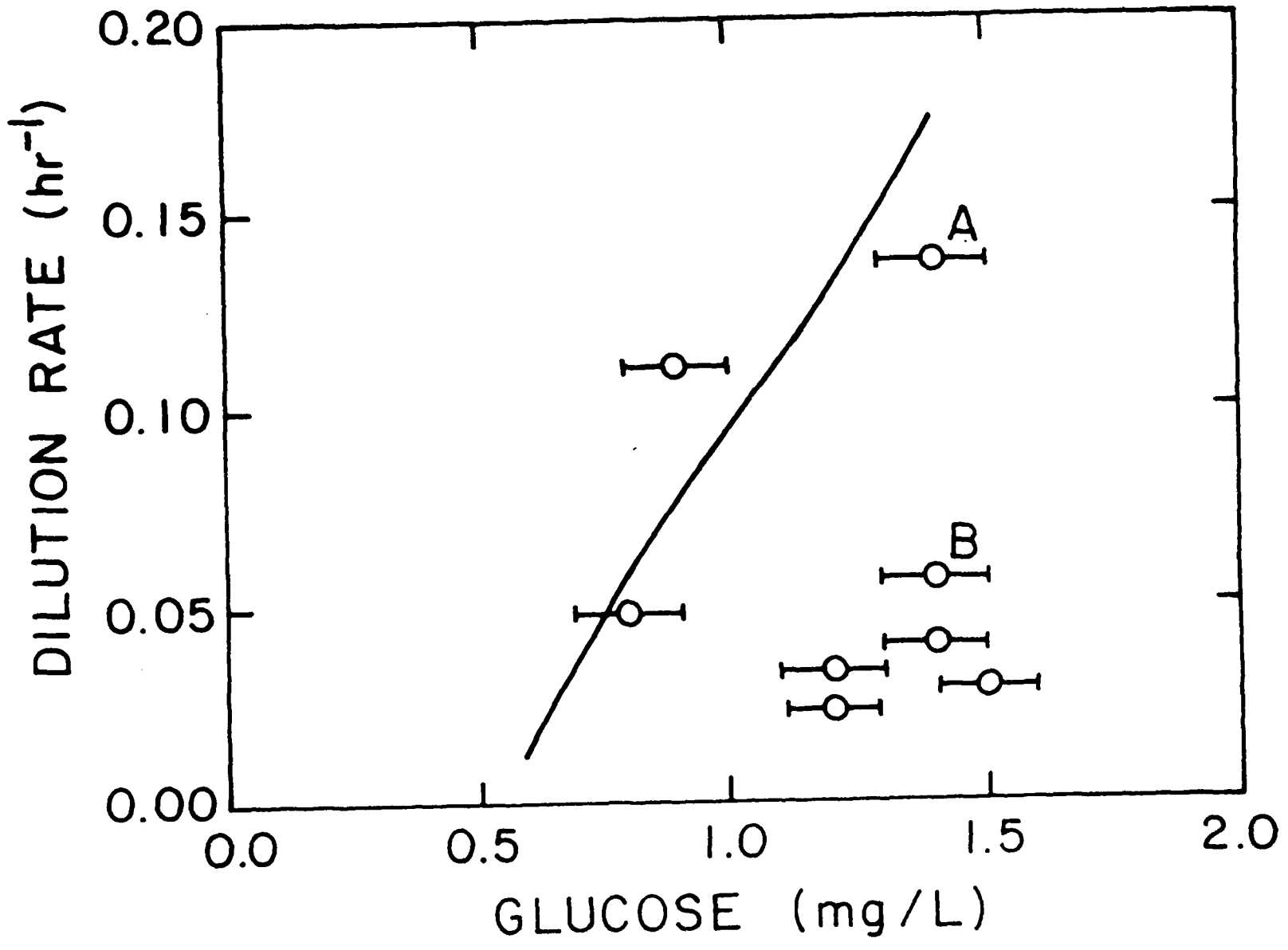
5. The plasmid copy number is shown as a function of growth rate for a derivative of a ColE1 type plasmid in *E. coli* B/r. The binding constant between RNA I and RNA II for this plasmid was 9.2×10^{-13} in the presence of the ROM protein. The solid line represents the case where RNA I and RNA II transcription rates vary with growth rate in a manner proportional to the changes in average transcription rate in the cell. The dashed line represents model predictions if RNA I and RNA II transcription rates are independent of growth rate and set at the observed values obtained under conditions of exponential growth in hatch culture.

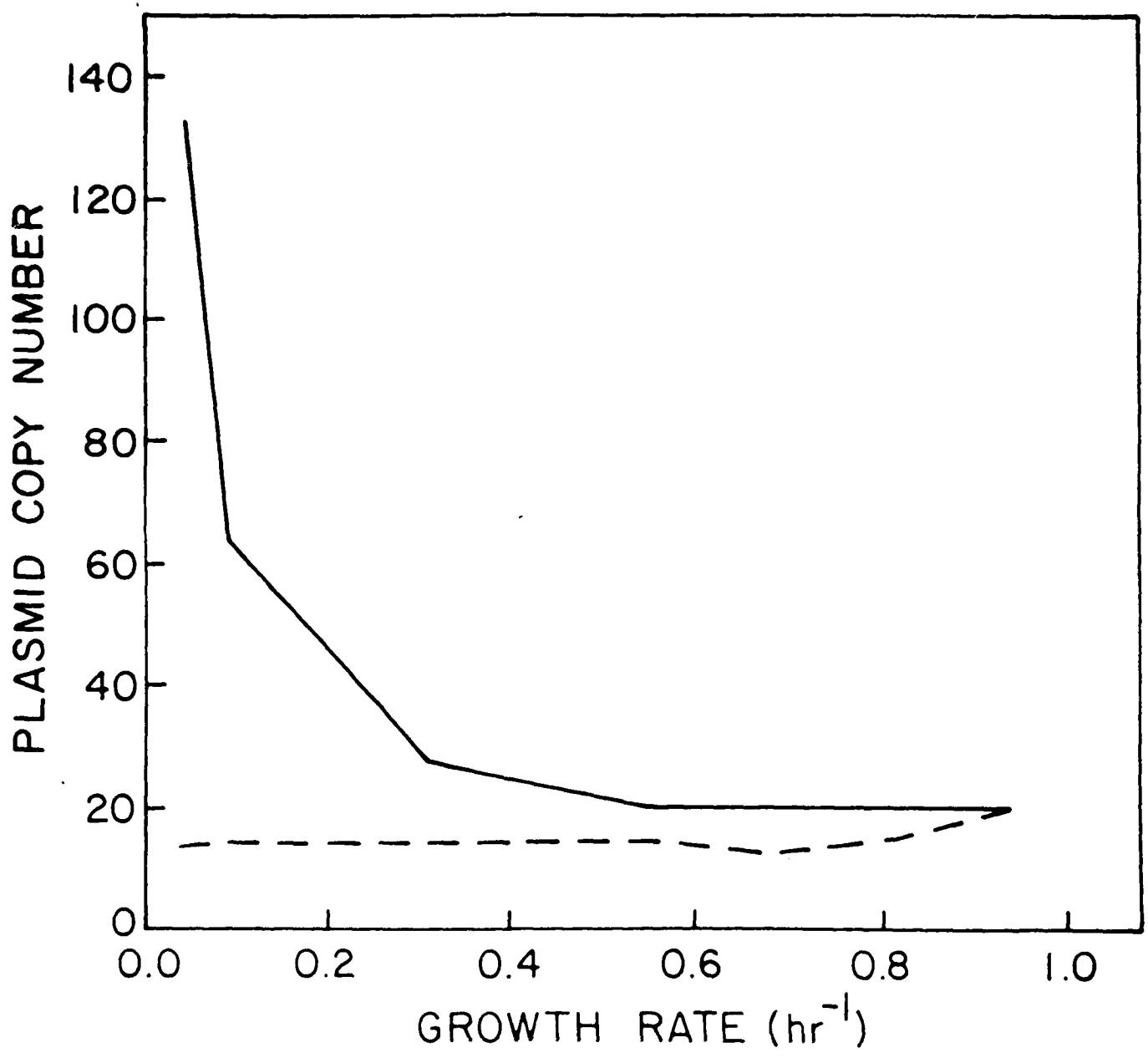
6. The prediction of plasmid content on growth rate for the same conditions as described in Figure 5.

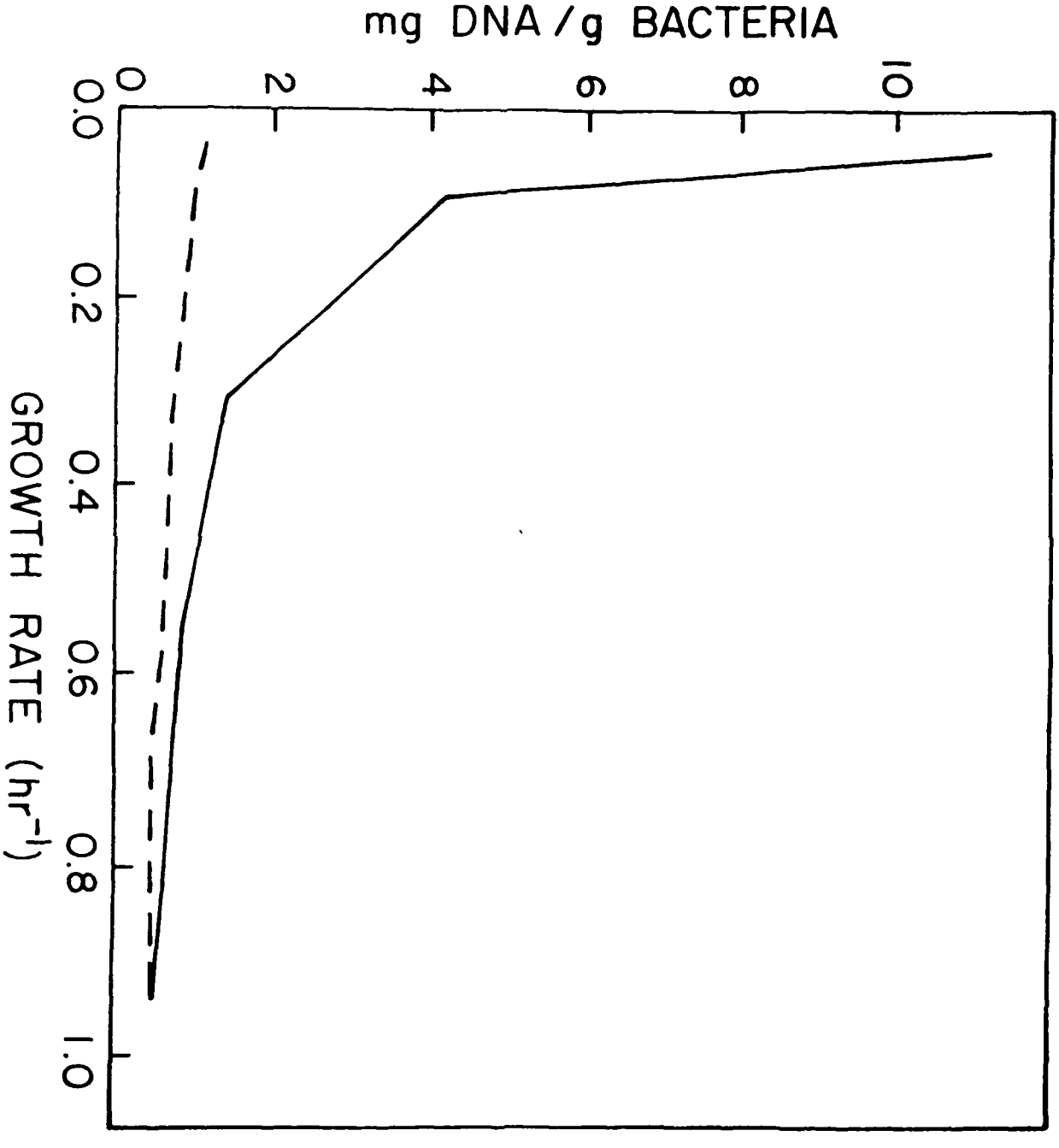

[INTERNAL GLUCOSE] (mg/L)
40
30
20
10
0
0.00
0.05
0.10
0.15
0.20
GROWTH RATE (1/hr)


1/GROWTH RATE (hr)
0
10
20
30
40
50
0.0
0.5
1.0
1.5
2.0
1/[SUBSTRATE] (1/mg/L)


50
40
30
20
10
0
0.0
0.5
1.0
1.5
2.0
1/GROWTH RATE (hr)
1/[SUBSTRATE] (1/mg/L)


0.20
0.15
0.10
0.05
0.00
DILUTION RATE (hr⁻¹)
A
B
0.0
0.5
1.0
1.5
2.0
GLUCOSE (mg/L)


PLASMID COPY NUMBER
140
120
100
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
GROWTH RATE (hr^{-1})


mg DNA / g BACTERIA
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
GROWTH RATE (hr⁻¹)

DISTRIBUTION LIST MOLECULAR BIOLOGY PROGRAM

ANNUAL, FINAL, AND TECHNICAL REPORTS (One copy each except as noted)

Dr. Lewis F. Affronti
George Washington University
Department of Microbiology
2300 I ST NW
Washington, DC 20037

Dr. J. Thomas August
The Johns Hopkins University
School of Medicine
720 Rutland Avenue
Baltimore, MD 21205

Dr. Myron L. Bender
Chemistry Department
Northwestern University
Evanston, IL 60201

Dr. R. P. Blakemore
University of New Hampshire
Department of Microbiology
Durham, New Hampshire 03824

Dr. Ronald Breslow
Columbia University
Department of Chemistry
New York, NY 10027

Dr. James P. Collman
Department of Chemistry
Stanford University
Stanford, California 94305

Dr. Alvin Crumbliss
North Carolina Biotechnology Center
Post Office Box 12235
Research Triangle Park, NC 27709

Dr. Marlene Deluca
University of California, San Diego
Department of Chemistry
La Jolla, CA 92093

Dr. Bruce Erickson
Chemistry Department
University of North Carolina
Chapel Hill, NC 27514

Dr. Richard B. Frankel
Massachusetts Institute of Technology
Francis Bitter National Laboratory
Cambridge, MA 02139

Dr. Hans Frauenfelder
Department of Physics
University of Illinois
Urbana, IL 61801

Dr. Bruce Gaber
Naval Research Laboratory
Code 6190
Washington, DC 20375

Dr. R. W. Giese
Northeastern Univ
Section of Medicinal Chemistry
360 Hungtington Ave
Boston, MA 02115

Dr. Barry Honig
Columbia University
Dept of Biochemistry and Molecular Biophysics
630 West 168th St.
New York, NY 10032

Dr. Alex Karu
Department of Plant Pathology
College of Natural Resources
University of California
Berkeley, CA 94720

Dr. Robert G. Kemp
University of Health Sciences
Chicago Medical school
Department of Biological Chemistry
3333 Green Bay Road
North Chicago, IL 60064

Dr. Ghobind H. Khorana
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Dr. Richard Laursen
Chemistry Department
Boston University
590 Commonwealth Avenue
Boston, MA 02215

Dr. Robert W. Lenz
Chemical Engineering Department
University of Massachusetts
Amherst, MA 01003

Dr. Harden M. McConnell
Stanford University
Department of Chemistry
Stanford, CA 94305

Dr. Kristin Bowman Mertes
University of Kansas
Department of Chemistry
Lawrence, Kansas 66045

Dr. Edgard F. Meyer
Texas A&M University
Department of Biochemistry and Biophysics
Box 3578
College Station, TX 77843

Dr. Jiri Novotny
Laboratory of Cellular and Molecular Research
Massachusetts General Hospital
Boston, MA 02114

Dr. Carl O. Pabo
Johns Hopkins Medical School
Department of Biophysics
Baltimore, MD 21205

Dr. Franklyn Prendergast
Mayo Foundation
200 First St. SW
Rochester, MN 55905

Dr. Naftali Primor
New York Zoological Society
New York Aquarium
Osborne Laboratory of Marine Science
Brooklyn, NY 11224

Dr. K. S. Rajan
Illinois Institute of Technology
Research Institute
10 W. 35th St.
Chicago, IL 60616

Dr. C. Patrick Reynolds
Naval Medical Research Institute
Transplantation Research Program Center
Bethesda, MD 20814

Dr. Alexander Rich
Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. J. H. Richards
California Institute of Technology
Division of Chemistry and Chemical Engineering
Pasadena, CA 91125

Dr. J. S. Richardson
Duke University School of Medicine
Department of Anatomy
Durham, NC 27910

Dr. Richard Roblin
Genex Corporation
16020 Industrial Drive
Gaithersburg, MD 20877

Dr. Peter G. Schultz
Department of Chemistry
University of California
Berkeley, CA 94720

Dr. Michael E. Selsted
Department of Medicine
UCLA School of Medicine
37-055 CHS
Los Angeles, CA 90024

Dr. Michael Shuler
School of Chemical Engineering
Cornell University
Ithaca, New York 14853

Dr. David S. Sigman
UCLA School of Medicine
Department of Biological Chemistry
Los Angeles, CA 90024

Dr. John M. Stewart
University of Colorado Health Science Center
Department of Biochemistry
Denver, CO 80262

Dr. Dan W. Urry
Laboratory of Molecular Biophysics
University of Alabama
P. O. Box 311
Birmingham, AL 35294

Dr. J. Herbert Waite
College of Marine Studies
University of Deleware
Lewes, DE 19958

Dr. Gerald D. Watt
Battelle-C. F. Kettering Research Laboratory
150 East South College Street
P. O. Box 268
Yellow Springs, Ohio 45387

Dr. Jon I Williams
Allied Corporation
Columbia Rd and Park Ave.
Morristown, NJ 07960

Dr. Eli D. Schmell, Code 1141MB
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Michael T. Marron, Code 1141MB
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Biochem 3/4

Dr. Margo G. Haygood
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Adminstrator (2 copies, Enclose DTIC Form 50)
Defense Technical Information Center
Building 5, Cameron Station
Alexandria, VA 22314

ANNUAL AND FINAL REPORTS ONLY (One copy each)
Commander
Chemical and Biological Sciences Division
Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709

Directorate of Life Sciences
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332

Chemistry and Atmospheric Sciences Directorate
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332

Director
Biotechnology Division
CRDEC
Aberdeen Proving Grounds, MD 21010-5423

Administrative Contracting Officer
ONR Resident Representative
(Address varies - obtain from your business office)

Director, Code 12
Applied Research and Technology Directorate
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Director, Code 22
Support Technology Directorate
Office of Naval Technology
800 North Quincy Street
Arlington, VA 22217-5000

Director, Code 112
Environmental Sciences Directorate
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Director, Code 113
Chemistry Division
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

FINAL AND TECHNICAL REPORTS ONLY
Director (6 copies)
Naval Research Laboratory
Attn: Technical Information Division, Code 2627
Washington, DC 20375